PUFA POLYKETIDE SYNTHASE SYSTEMS AND USES THEREOF

Cross-Reference to Related Applications

This application claims the benefit of priority under 35 U.S.C. § 119(e) from U.S. Provisional Application Serial No. 60/457,979, filed March 26, 2003, entitled "Modification of a *Schizochytrium* PKS System to Facilitate Production of Lipids Rich in Polyunsaturated Fatty Acids". This application is also a continuation-in-part of U.S. Patent Application Serial No. 10/124,800, filed April 16, 2002, which Claims the benefit of priority under 35 U.S.C. § 119(e) to: U.S. Provisional Application Serial No. 60/284,066, filed April 16, 2001; U.S. Provisional Application Serial No. 60/298,796, filed June 15, 2001; and U.S. Provisional Application Serial No. 60/323,269, filed September 18, 2001. U.S. Patent Application Serial No. 10/124,800, *supra*, is also a continuation-in-part of U.S. Application Serial No. 09/231,899, filed January 14, 1999, now U.S. Patent No. 6,566,583. Each of the above-identified patent applications is incorporated herein by reference in its entirety.

This application does not claim the benefit of priority from U.S. Application Serial No. 09/090,793, filed June 4, 1998, now U.S. Patent No. 6,140,486, although U.S. Application Serial No. 09/090,793 is incorporated herein by reference in its entirety.

Field of the Invention

This invention relates to polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems from microorganisms, including eukaryotic organisms, such as Thraustochytrid microorganisms. More particularly, this invention relates to nucleic acids encoding non-bacterial PUFA PKS systems, to non-bacterial PUFA PKS systems, to genetically modified organisms comprising non-bacterial PUFA PKS systems, and to methods of making and using the non-bacterial PUFA PKS systems disclosed herein. This invention also relates to genetically modified microorganisms and methods to efficiently produce lipids (triacylglyerols (TAG), as well as membrane-associated phospholipids (PL)) enriched in various polyunsaturated fatty acids (PUFAs) and particularly, eicosapentaenoic acid (C20:5, ω-3; EPA) by manipulation of a PUFA polyketide synthase (PKS) system.

Background of the Invention

Polyketide synthase (PKS) systems are generally known in the art as enzyme complexes derived from fatty acid synthase (FAS) systems, but which are often highly modified to produce specialized products that typically show little resemblance to fatty acids. It has now been shown, however, that polyketide synthase systems exist in marine bacteria and certain microalgae that are capable of synthesizing PUFAs from malonyl-CoA. The PKS pathways for PUFA synthesis in *Shewanella* and another marine bacteria, *Vibrio marinus*, are described in detail in U.S. Patent No. 6,140,486. The PKS pathways for PUFA synthesis in the eukaryotic Thraustochytrid, *Schizochytrium* is described in detail in U.S. Patent 6,566,583. Finally, the PKS pathways for PUFA synthesis in eukaryotes such as members of Thraustochytriales, including the complete structural description of the PUFA PKS pathway in *Schizochytrium* and the identification of the PUFA PKS pathway in *Thraustochytrium*, including details regarding uses of these pathways, are described in detail in U.S. Patent Application Publication No. 20020194641, published December 19, 2002 (corresponding to U.S. Patent Application Serial No. 10/124,800, filed April 16, 2002).

Researchers have attempted to exploit polyketide synthase (PKS) systems that have been described in the literature as falling into one of three basic types, typically referred to as: Type II, Type I and modular. The Type II system is characterized by separable proteins, each of which carries out a distinct enzymatic reaction. The enzymes work in concert to produce the end product and each individual enzyme of the system typically participates several times in the production of the end product. This type of system operates in a manner analogous to the fatty acid synthase (FAS) systems found in plants and bacteria. Type I PKS systems are similar to the Type II system in that the enzymes are used in an iterative fashion to produce the end product. The Type I differs from Type II in that enzymatic activities, instead of being associated with separable proteins, occur as domains of larger proteins. This system is analogous to the Type I FAS systems found in animals and fungi.

In contrast to the Type I and II systems, in modular PKS systems, each enzyme domain is used only once in the production of the end product. The domains are found in very large proteins and the product of each reaction is passed on to another domain in the

PKS protein. Additionally, in all of the PKS systems described above, if a carbon-carbon double bond is incorporated into the end product, it is always in the *trans* configuration.

In the Type I and Type II PKS systems described above, the same set of reactions is carried out in each cycle until the end product is obtained. There is no allowance for the introduction of unique reactions during the biosynthetic procedure. The modular PKS systems require huge proteins that do not utilize the economy of iterative reactions (i.e., a distinct domain is required for each reaction). Additionally, as stated above, carbon-carbon double bonds are introduced in the *trans* configuration in all of the previously described PKS systems.

Polyunsaturated fatty acids (PUFAs) are critical components of membrane lipids in most eukaryotes (Lauritzen et al., *Prog. Lipid Res.* **40** 1 (2001); McConn et al., *Plant J.* **15**, 521 (1998)) and are precursors of certain hormones and signaling molecules (Heller et al., *Drugs* **55**, 487 (1998); Creelman et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 355 (1997)). Known pathways of PUFA synthesis involve the processing of saturated 16:0 or 18:0 fatty acids (the abbreviation X:Y indicates an acyl group containing X carbon atoms and Y double bonds (usually *cis* in PUFAs); double-bond positions of PUFAs are indicated relative to the methyl carbon of the fatty acid chain (ω3 or ω6) with systematic methylene interruption of the double bonds) derived from fatty acid synthase (FAS) by elongation and aerobic desaturation reactions (Sprecher, *Curr. Opin. Clin. Nutr. Metab. Care* **2**, 135 (1999); Parker-Barnes et al., *Proc. Natl. Acad. Sci. USA* **97**, 8284 (2000); Shanklin et al., *Annu. Rev. Plant Physiol. Plant Nol. Biol.* **49**, 611 (1998)). Starting from acetyl-CoA, the synthesis of docosahexaenoic acid (DHA) requires approximately 30 distinct enzyme activities and nearly 70 reactions including the four repetitive steps of the fatty acid synthesis cycle. Polyketide synthases (PKSs) carry out some of the same reactions as FAS (Hopwood et al., *Annu. Rev. Genet.* **24**, 37 (1990); Bentley et al., *Annu. Rev. Microbiol.* **53**, 411 (1999)) and use the same small protein (or domain), acyl carrier protein (ACP), as a covalent attachment site for the growing carbon chain. However, in these enzyme systems, the complete cycle of reduction, dehydration and reduction seen in FAS is often abbreviated so that a highly derivatized carbon chain is produced, typically containing many keto- and hydroxy-groups as well as

carbon-carbon double bonds in the *trans* configuration. The linear products of PKSs are often cyclized to form complex biochemicals that include antibiotics and many other secondary products (Hopwood et al., (1990) *supra*; Bentley et al., (1999), *supra*; Keating et al., *Curr. Opin. Chem. Biol.* **3**, 598 (1999)).

Very long chain PUFAs such as docosahexaenoic acid (DHA; 22:6ω3) and eicosapentaenoic acid (EPA; 20:5ω3) have been reported from several species of marine bacteria, including *Shewanella* sp (Nichols et al., *Curr. Op. Biotechnol.* **10**, 240 (1999); Yazawa, *Lipids* **31**, S (1996); DeLong et al., *Appl. Environ. Microbiol.* **51**, 730 (1986)). Analysis of a genomic fragment (cloned as plasmid pEPA) from *Shewanella* sp. strain SCRC2738 led to the identification of five open reading frames (Orfs), totaling 20 Kb, that are necessary and sufficient for EPA production in *E. coli* (Yazawa, (1996), *supra*). Several of the predicted protein domains were homologues of FAS enzymes, while other regions showed no homology to proteins of known function. At least 11 regions within the five Orfs were identifiable as putative enzyme domains (See Metz et al., *Science* 293:290-293 (2001)). When compared with sequences in the gene databases, seven of these were more strongly related to PKS proteins than to FAS proteins. Included in this group were domains putatively encoding malonyl-CoA:ACP acyltransferase (MAT), β-ketoacyl-ACP synthase (KS), β-ketoacyl-ACP reductase (KR), acyltransferase (AT), phosphopantetheine transferase, chain length (or chain initiation) factor (CLF) and a highly unusual cluster of six ACP domains (i.e., the presence of more than two clustered ACP domains had not previously been reported in PKS or FAS sequences). It is likely that the PKS pathway for PUFA synthesis that has been identified in *Shewanella* is widespread in marine bacteria. Genes with high homology to the *Shewanella* gene cluster have been identified in *Photobacterium profundum* (Allen et al., *Appli. Environ. Microbiol.* 65:1710 (1999)) and in *Moritella marina* (*Vibrio marinus*) (see U.S. Patent No. 6,140,486, *ibid.*, and Tanaka et al., *Biotechnol. Lett.* 21:939 (1999)).

Polyunsaturated fatty acids (PUFAs) are considered to be useful for nutritional, pharmaceutical, industrial, and other purposes. An expansive supply of PUFAs from natural sources and from chemical synthesis are not sufficient for commercial needs. A major

current source for PUFAs is from marine fish; however, fish stocks are declining, and this may not be a sustainable resource. Additionally, contamination, both heavy metal and toxic organic molecules, is a serious issue with oil derived from marine fish. Vegetable oils derived from oil seed crops are relatively inexpensive and do not have the contamination issues associated with fish oils. However, the PUFAs found in commercially developed plant oils are typically limited to linoleic acid (eighteen carbons with 2 double bonds, in the delta 9 and 12 positions - 18:2 delta 9,12) and linolenic acid (18:3 delta 9,12,15). In the conventional pathway for PUFA synthesis, medium chain-length saturated fatty acids (products of a fatty acid synthase (FAS) system) are modified by a series of elongation and desaturation reactions. Because a number of separate desaturase and elongase enzymes are required for fatty acid synthesis from linoleic and linolenic acids to produce the more saturated and longer chain PUFAs, engineering plant host cells for the expression of PUFAs such as EPA and docosahexaenoic acid (DHA) may require expression of several separate enzymes to achieve synthesis. Additionally, for production of useable quantities of such PUFAs, additional engineering efforts may be required, for example, engineering the down regulation of enzymes that compete for substrate, engineering of higher enzyme activities such as by mutagenesis or targeting of enzymes to plastid organelles. Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated material alone or in combination in a heterologous system which can be manipulated to allow production of commercial quantities of PUFAs.

The discovery of a PUFA PKS system in marine bacteria such as *Shewanella* and *Vibrio marinus* (see U.S. Patent No. 6,140,486, *ibid.*) provides a resource for new methods of commercial PUFA production. However, these marine bacteria have limitations which may ultimately restrict their usefulness on a commercial level. First, although U.S. Patent No. 6,140,486 discloses that these marine bacteria PUFA PKS systems can be used to genetically modify plants, the marine bacteria naturally live and grow in cold marine environments and the enzyme systems of these bacteria do not function well above 22°C. In contrast, many crop plants, which are attractive targets for genetic manipulation using the

PUFA PKS system, have normal growth conditions at temperatures above 22°C and ranging to higher than 40°C. Therefore, the PUFA PKS systems from these marine bacteria are not predicted to be readily adaptable to plant expression under normal growth conditions. Additionally, the known marine bacteria PUFA PKS systems do not directly produce triacylglyerols (TAG), whereas direct production of TAG would be desirable because TAG are a lipid storage product, and as a result, can be accumulated at very high levels in cells, as opposed to a "structural" lipid product (e.g. phospholipids), which can generally only accumulate at low levels.

With regard to the production of eicosapentaenoic acid (EPA) in particular, researchers have tried to produce EPA with microbes by growing them in both photosynthetic and heterotrophic cultures. They have also used both classical and directed genetic approaches in attempts to increase the productively of the organisms under culture conditions. Other researchers have attempted to produce EPA in oil-seed crop plants by introduction of genes encoding various desaturase and elongase enzymes.

Researchers have attempted to use cultures of red microalgae (Monodus), diatoms (e.g. *Phaeodactylum*), other microalgae and fungi (e.g. *Mortierella* cultivated at low temperatures). However, in all cases, productivity was low compared to existing commercial microbial production systems for other long chain PUFAs such as DHA. In many cases, the EPA occurred primarily in the phospholipids (PL) rather than the triacylglycerols (TAG). Since productivity of microalgae under heterotrophic growth conditions can be much higher than under phototrophic conditions, researchers have attempted, and achieved, trophic conversion by introduction of genes encoding specific sugar transporters. However, even with the newly acquired heterotrophic capability, productivity in terms of oil remained relatively low.

Efforts to produce EPA in oil-seed crop plants by modification of the endogenous fatty acid biosynthesis pathway have only yielded plants with very low levels of the PUFA in their oils. As discussed above, several marine bacteria have been shown to produce PUFAs (EPA as well as DHA). However, these bacteria do not produce TAG and the EPA is found primarily in the PL membranes. The levels of EPA produced as well as the growth

characteristics of these particular marine bacteria (discussed above) limit their utility for commercial production of EPA.

Therefore, there is a need in the art for other PUFA PKS systems having greater flexibility for commercial use, and for a biological system that efficiently produces quantities of lipids (PL and TAG) enriched in desired PUFAs, such as EPA, in a commercially useful production process.

Summary of the Invention

One embodiment of the present invention relates to an isolated nucleic acid molecule. The nucleic acid molecule comprises a nucleic acid sequence selected from: (a) a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) a nucleic acid sequence encoding an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) a nucleic acid sequence encoding an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) a nucleic acid sequence encoding an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID

NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (e) a nucleic acid sequence encoding an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and/or (f) a nucleic acid sequence that is fully complementary to the nucleic acid sequence of (a), (b), (c), (d), or (e). In one aspect, the nucleic acid sequence encodes an amino acid sequence selected from: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, and biologically active fragments thereof. In one aspect, the nucleic acid sequence is selected from the group consisting of: SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, and SEQ ID NO:67.

Another embodiment of the present invention relates to a recombinant nucleic acid molecule comprising any of the above-described nucleic acid molecules, operatively linked to at least one transcription control sequence.

Yet another embodiment of the present invention relates to a recombinant cell transfected with any of the above-described recombinant nucleic acid molecules.

Another embodiment of the present invention relates to a genetically modified microorganism, wherein the microorganism expresses a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, wherein the at least one domain of the PUFA PKS system comprises an amino acid sequence selected from: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active

fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and/or (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. The microorganism is genetically modified to affect the activity of the PKS system.

In one aspect, the microorganism is genetically modified by transfection with a recombinant nucleic acid molecule encoding the at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. For example, the microorganism can include a Thraustochytrid, such as a *Schizochytrium*. In one aspect, such a microorganism has been further genetically modified to recombinantly express at least one nucleic acid molecule encoding at least one biologically active domain from a PKS system selected from the group consisting of: a bacterial PUFA PKS system, a Type I PKS system, a Type II PKS system, a modular PKS system, and a non-bacterial PUFA PKS system. The non-bacterial

PUFA PKS system can include a Thraustochytrid PUFA PKS system and in one aspect, a *Schizochytrium* PUFA PKS system.

In another aspect, the microorganism endogenously expresses a PKS system comprising the at least one domain of the PUFA PKS system, and wherein the genetic modification is in a nucleic acid sequence encoding at least one domain of the PUFA PKS system. In another aspect, such a microorganism has been further genetically modified to recombinantly express at least one nucleic acid molecule encoding at least one biologically active domain from a PKS system selected from the group consisting of: a bacterial PUFA PKS system, a Type I PKS system, a Type II PKS system, a modular PKS system, and a non-bacterial PUFA PKS system (e.g., a Thraustochytrid PUFA PKS system, such as a *Schizochytrium* PUFA PKS system).

In another aspect, the microorganism endogenously expresses a PUFA PKS system comprising the at least one biologically active domain of a PUFA PKS system, and wherein the genetic modification comprises expression of a recombinant nucleic acid molecule selected from the group consisting of a recombinant nucleic acid molecule encoding at least one biologically active domain from a second PKS system and a recombinant nucleic acid molecule encoding a protein that affects the activity of the endogenous PUFA PKS system. The biologically active domain from a second PKS system can include, but is not limited to: (a) a domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system from a Thraustochytrid microorganism; (b) a domain of a PUFA PKS system from a microorganism identified by the following method: (i) selecting a microorganism that produces at least one PUFA; and, (ii) identifying a microorganism from (i) that has an ability to produce increased PUFAs under dissolved oxygen conditions of less than about 5% of saturation in the fermentation medium, as compared to production of PUFAs by the microorganism under dissolved oxygen conditions of greater than about 5% of saturation in the fermentation medium; (c) a domain comprising an amino acid sequence selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, and biologically

active fragments thereof; and (d) a domain comprising an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to the amino acid sequence of (c), wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. In one aspect, recombinant nucleic acid molecule encodes a phosphopantetheine transferase. In one aspect, the second PKS system is selected from the group consisting of: a bacterial PUFA PKS system, a type I PKS system, a type II PKS system, a modular PKS system, and a non-bacterial PUFA PKS system (e.g., a eukaryotic PUFA PKS system, such as a Thraustochytrid PUFA PKS system, including, but not limited to a *Schizochytrium* PUFA PKS system).

Yet another embodiment of the present invention relates to a genetically modified plant, wherein the plant has been genetically modified to recombinantly express a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, wherein the domain comprises an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid

sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and/or (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. In one aspect, the at least one domain of the PUFA PKS system comprises an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 and SEQ ID NO:68 and biologically active fragments thereof. In one aspect, the plant has been further genetically modified to recombinantly express at least one nucleic acid molecule encoding at least one biologically active domain from a PKS system selected from the group consisting of: a bacterial PUFA PKS system, a Type I PKS system, a Type II PKS system, a modular PKS system, and a non-bacterial PUFA PKS system (e.g., a Thraustochytrid PUFA PKS system, such as a *Schizochytrium* PUFA PKS system).

Yet another embodiment of the present invention relates to a method to produce a bioactive molecule that is produced by a polyketide synthase system, comprising culturing under conditions effective to produce the bioactive molecule a genetically modified organism that expresses a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, wherein the at least one domain of the PUFA PKS system comprises an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60,

SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and/or (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system.

In one aspect, the organism endogenously expresses a PKS system comprising the at least one domain of the PUFA PKS system, and wherein the genetic modification is in a nucleic acid sequence encoding the at least one domain of the PUFA PKS system. In one aspect, the genetic modification changes at least one product produced by the endogenous PKS system, as compared to an organism wherein the PUFA PKS system has not been genetically modified.

In another aspect, the organism endogenously expresses a PKS system comprising the at least one biologically active domain of the PUFA PKS system, and the genetic

modification comprises transfection of the organism with a recombinant nucleic acid molecule selected from the group consisting of: a recombinant nucleic acid molecule encoding at least one biologically active domain from a second PKS system and a recombinant nucleic acid molecule encoding a protein that affects the activity of the PUFA PKS system. In one aspect, the genetic modification changes at least one product produced by the endogenous PKS system, as compared to an organism that has not been genetically modified to affect PUFA production.

In another aspect, the organism is genetically modified by transfection with a recombinant nucleic acid molecule encoding the at least one domain of the polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system.

In another aspect, the organism produces a polyunsaturated fatty acid (PUFA) profile that differs from the naturally occurring organism without a genetic modification.

In another aspect, the organism endogenously expresses a non-bacterial PUFA PKS system, and wherein the genetic modification comprises substitution of a domain from a different PKS system for a nucleic acid sequence encoding at least one domain of the non-bacterial PUFA PKS system.

In yet another aspect, the organism endogenously expresses a non-bacterial PUFA PKS system that has been modified by transfecting the organism with a recombinant nucleic acid molecule encoding a protein that regulates the chain length of fatty acids produced by the PUFA PKS system.

In another aspect, the bioactive molecule is selected from: an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and/or a cholesterol lowering formulation. In one aspect, the bioactive molecule is an antibiotic. In another aspect, the bioactive molecule is a polyunsaturated fatty acid (PUFA). In yet another aspect, the bioactive molecule is a molecule including carbon-carbon double bonds in the *cis* configuration. In one aspect, the bioactive molecule is a molecule including

a double bond at every third carbon. In one aspect, the organism is a microorganism. In another aspect, the organism is a plant.

Another embodiment of the present invention relates to a method to produce a plant that has a polyunsaturated fatty acid (PUFA) profile that differs from the naturally occurring plant, comprising genetically modifying cells of the plant to express a PKS system comprising at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system, wherein the at least one domain of the PUFA PKS system comprises an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90%

identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system.

Another embodiment of the present invention relates to a method to modify an endproduct containing at least one fatty acid, comprising adding to the endproduct an oil produced by a recombinant host cell that expresses at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system, wherein the at least one domain of a PUFA PKS system comprises an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide

synthase (PKS) system; and (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. In one aspect, the endproduct is selected from: a dietary supplement, a food product, a pharmaceutical formulation, a humanized animal milk, and an infant formula. In one aspect, the pharmaceutical formulation is selected from the group consisting of an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and a cholesterol lowering formulation. In one aspect, the endproduct is used to treat a condition selected from the group consisting of: chronic inflammation, acute inflammation, gastrointestinal disorder, cancer, cachexia, cardiac restenosis, neurodegenerative disorder, degenerative disorder of the liver, blood lipid disorder, osteoporosis, osteoarthritis, autoimmune disease, preeclampsia, preterm birth, age related maculopathy, pulmonary disorder, and peroxisomal disorder.

Yet another embodiment of the present invention relates to a method to produce a humanized animal milk, comprising genetically modifying milk-producing cells of a milk-producing animal with at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system, wherein the at least one domain of the PUFA PKS system comprises an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ

ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (c) an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to SEQ ID NO:54, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; (d) an amino acid sequence that is at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:60, SEQ ID NO:62 and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and (e) an amino acid sequence that is at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system.

Another embodiment of the present invention relates to a genetically modified Thraustochytrid microorganism, wherein the microorganism has an endogenous polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, and wherein the endogenous PUFA PKS system has been genetically modified to alter the expression profile of a polyunsaturated fatty acid (PUFA) by the Thraustochytrid microorganism as compared to the Thraustochytrid microorganism in the absence of the genetic modification.

In one aspect, the endogenous PUFA PKS system has been modified by mutagenesis of a nucleic acid sequence that encodes at least one domain of the endogenous PUFA PKS system. In one aspect, the modification is produced by targeted mutagenesis. In another aspect, the modification is produced by classical mutagenesis and screening.

In another aspect, the endogenous PUFA PKS system has been modified by deleting at least one nucleic acid sequence that encodes at least one domain of the endogenous PUFA

PKS system and inserting therefore a nucleic acid sequence encoding a homologue of the endogenous domain to alter the PUFA production profile of the Thraustochytrid microorganism, wherein the homologue has a biological activity of at least one domain of a PKS system. In one aspect, the homologue of the endogenous domain comprises a modification, as compared to the endogenous domain, selected from the group consisting of at least one deletion, insertion or substitution that results in an alteration of PUFA production profile by the microorganism. In another aspect, the amino acid sequence of the homologue is at least about 60% identical, and more preferably about 70% identical, and more preferably about 80% identical, and more preferably about 90% identical to the amino acid sequence of the endogenous domain. In one aspect, homologue of the endogenous domain is a domain from a PUFA PKS system of another Thraustochytrid microorganism.

In another aspect, the endogenous PUFA PKS system has been modified by deleting at least one nucleic acid sequence that encodes at least one domain of the endogenous PUFA PKS system and inserting therefore a nucleic acid sequence encoding at least one domain of a PKS system from a different microorganism. In one aspect, the nucleic acid sequence encoding at least one domain of a PKS system from a different microorganism is from a bacterial PUFA PKS system. For example, the different microorganism can be a marine bacteria having a PUFA PKS system that naturally produces PUFAs at a temperature of about 25°C or greater. In one aspect, the marine bacteria is selected from the group consisting of *Shewanella olleyana* and *Shewanella japonica*. In one aspect, the domain of a PKS system from a different microorganism is from a PKS system selected from the group consisting of: a Type I PKS system, a Type II PKS system, a modular PKS system, and a PUFA PKS system from a different Thraustochytrid microorganism.

In any of the above aspects, the domain of the endogenous PUFA PKS system can include, but is not limited to, a domain having a biological activity of at least one of the following proteins: malonyl-CoA:ACP acyltransferase (MAT), β-keto acyl-ACP synthase (KS), ketoreductase (KR), acyltransferase (AT), FabA-like β-hydroxy acyl-ACP dehydrase (DH), phosphopantetheine transferase, chain length factor (CLF), acyl carrier protein (ACP), enoyl ACP-reductase (ER), an enzyme that catalyzes the synthesis of *trans*-2-acyl-ACP, an

enzyme that catalyzes the reversible isomerization of *trans*-2-acyl-ACP to *cis*-3-acyl-ACP, and an enzyme that catalyzes the elongation of *cis*-3-acyl-ACP to *cis*-5-β-keto-acyl-ACP. In any of the above aspects, the domain of the endogenous PUFA PKS system can include an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; and (b) an amino acid sequence that is at least about 60% identical, and more preferably at least about 70% identical, and more preferably at least about 80% identical, and more preferably at least about 90% identical, to an amino acid sequence of (a), wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system.

In one aspect, the PUFA production profile is altered to initiate, increase or decrease production of eicosapentaenoic acid (EPA) by the microorganism. In another aspect, the PUFA production profile is altered to initiate, increase or decrease production of docosahexaenoic acid (DHA) by the microorganism. In another aspect, the PUFA production profile is altered to initiate, increase or decrease production of one or both isomers of docosapentaenoic acid (DPA) by the microorganism. In another aspect, the PUFA production profile is altered to initiate, increase or decrease production of arachidonic acid (ARA) by the microorganism. In another aspect, the Thraustochytrid is from a genus selected from the group consisting of *Schizochytrium, Thraustochytrium*, and *Japonochytrium*. In another aspect, the Thraustochytrid is from the genus *Schizochytrium*. In another aspect, the Thraustochytrid is from a *Schizochytrium* species selected from the group consisting of: *Schizochytrium aggregatum*, *Schizochytrium limacinum*, and *Schizochytrium minutum*. In another aspect, the Thraustochytrid is from the genus *Thraustochytrium*.

Yet another embodiment of the present invention relates to a genetically modified *Schizochytrium* that produces eicosapentaenoic acid (EPA), wherein the *Schizochytrium* has an endogenous polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system comprising a genetic modification in at least one nucleic acid sequence that encodes at least one domain of the endogenous PUFA PKS system that results in the production of EPA by the *Schizochytrium*. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding at least one domain having a biological activity of at least one of the following proteins: malonyl-CoA:ACP acyltransferase (MAT), β-keto acyl-ACP synthase (KS), ketoreductase (KR), acyltransferase (AT), FabA-like β-hydroxy acyl-ACP dehydrase (DH), phosphopantetheine transferase, chain length factor (CLF), acyl carrier protein (ACP), enoyl ACP-reductase (ER), an enzyme that catalyzes the synthesis of *trans*-2-acyl-ACP, an enzyme that catalyzes the reversible isomerization of *trans*-2-acyl-ACP to *cis*-3-acyl-ACP, and an enzyme that catalyzes the elongation of *cis*-3-acyl-ACP to *cis*-5-β-keto-acyl-ACP. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding at least one domain from the open reading frame encoding SEQ ID NO:2 of the endogenous PUFA PKS system. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding at least one domain from the open reading frame encoding SEQ ID NO:4 of the endogenous PUFA PKS system. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding at least one domain from the open reading frame encoding SEQ ID NO:6 of the endogenous PUFA PKS system. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding at least one domain having a biological activity of at least one of the following proteins: β-keto acyl-ACP synthase (KS), FabA-like β-hydroxy acyl-ACP dehydrase (DH), chain length factor (CLF), an enzyme that catalyzes the synthesis of *trans*-2-acyl-ACP, an enzyme that catalyzes the reversible isomerization of *trans*-2-acyl-ACP to *cis*-3-acyl-ACP, and an enzyme that catalyzes the elongation of *cis*-3-acyl-ACP to *cis*-5-β-keto-acyl-ACP. In one aspect, the *Schizochytrium* comprises a genetic modification in at least one nucleic acid sequence encoding an amino acid sequence selected from the group consisting

of SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28 and SEQ ID NO:30 of the endogenous PUFA PKS system. In one aspect, the *Schizochytrium* has been modified by deleting at least one nucleic acid sequence that encodes at least one domain of the endogenous PUFA PKS system and inserting therefore a nucleic acid sequence encoding at least one domain of a PKS system from a non-*Schizochytrium* microorganism. In one aspect, the non-*Schizochytrium* microorganism grows and produces PUFAs at temperature of at least about 15°C, and more preferably at least about 20°C, and more preferably at least about 25°C, and more preferably at least about 30°C, and more preferably between about 20°C and about 40°C. In one aspect, the nucleic acid sequence encoding at least one domain of a PKS system from a non-*Schizochytrium* microorganism is from a bacterial PUFA PKS system. In one aspect, the bacterial PUFA PKS system is from a bacterium selected from the group consisting of *Shewanella olleyana* and *Shewanella japonica*. In another aspect, the nucleic acid sequence encoding at least one domain of a PKS system is selected from the group consisting of a Type I PKS system, a Type II PKS system, a modular PKS system, and a non-bacterial PUFA PKS system (e.g., a eukaryotic PUFA PKS system, such as a Thraustochytrid PUFA PKS system).

Another embodiment of the present invention relates to a genetically modified *Schizochytrium* that produces increased amounts of docosahexaenoic acid (DHA) as compared to a non-genetically modified *Schizochytrium*, wherein the *Schizochytrium* has an endogenous polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system comprising a genetic modification in at least one nucleic sequence that encodes at least one domain of the endogenous PUFA PKS system that results in increased the production of DHA by the *Schizochytrium*. In one aspect, at least one domain of the endogenous PUFA PKS system has been modified by substitution for at least one domain of a PUFA PKS system from *Thraustochytrium*. In one aspect, the ratio of DHA to DPA produced by the *Schizochytrium* is increased as compared to a non-genetically modified *Schizochytrium*.

Another embodiment of the present invention relates to a method to produce lipids enriched for at least one selected polyunsaturated fatty acid (PUFA), comprising culturing under conditions effective to produce the lipids a genetically modified Thraustochytrid

microorganism as described above or a genetically modified *Schizochytrium* as described above. In one aspect, the selected PUFA is eicosapentaenoic acid (EPA).

Yet another embodiment of the present invention relates to a method to produce eicosapentaenoic acid (EPA)-enriched lipids, comprising culturing under conditions effective to produce the EPA-enriched lipids a genetically modified Thraustochytrid microorganism, wherein the microorganism has an endogenous polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, and wherein the endogenous PUFA PKS system has been genetically modified in at least one domain to initiate or increase the production of EPA in the lipids of the microorganism as compared to in the absence of the modification.

Brief Description of the Figures

Fig. 1 is a graphical representation of the domain structure of the *Schizochytrium* PUFA PKS system.

Fig. 2 shows a comparison of domains of PUFA PKS systems from *Schizochytrium* and *Shewanella*.

Fig. 3 shows a comparison of domains of PUFA PKS systems from *Schizochytrium* and a related PKS system from *Nostoc* whose product is a long chain fatty acid that does not contain any double bonds.

Detailed Description of the Invention

The present invention generally relates to polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems, to genetically modified organisms comprising such PUFA PKS systems, to methods of making and using such systems for the production of products of interest, including bioactive molecules and particularly, PUFAs, such as DHA, DPA and EPA. As used herein, a PUFA PKS system generally has the following identifying features: (1) it produces PUFAs as a natural product of the system; and (2) it comprises several multifunctional proteins assembled into a complex that conducts both *iterative* processing of the fatty acid chain as well *non-iterative* processing, including *trans-cis* isomerization and enoyl reduction reactions in selected cycles (See Fig. 1, for example).

Reference to a PUFA PKS system refers collectively to all of the genes and their encoded products that work in a complex to produce PUFAs in an organism. Therefore, the PUFA PKS system refers specifically to a PKS system for which the natural products are PUFAs.

More specifically, first, a PUFA PKS system that forms the basis of this invention produces polyunsaturated fatty acids (PUFAs) as products (i.e., an organism that endogenously (naturally) contains such a PKS system makes PUFAs using this system). The PUFAs referred to herein are preferably polyunsaturated fatty acids with a carbon chain length of at least 16 carbons, and more preferably at least 18 carbons, and more preferably at least 20 carbons, and more preferably 22 or more carbons, with at least 3 or more double bonds, and preferably 4 or more, and more preferably 5 or more, and even more preferably 6 or more double bonds, wherein all double bonds are in the *cis* configuration. It is an object of the present invention to find or create via genetic manipulation or manipulation of the endproduct, PKS systems which produce polyunsaturated fatty acids of desired chain length and with desired numbers of double bonds. Examples of PUFAs include, but are not limited to, DHA (docosahexaenoic acid (C22:6, ω-3)), ARA (eicosatetraenoic acid or arachidonic acid (C20:4, n-6)), DPA (docosapentaenoic acid (C22:5, ω-6 or ω-3)), and EPA (eicosapentaenoic acid (C20:5, ω-3)).

Second, the PUFA PKS system described herein incorporates both iterative and non-iterative reactions, which distinguish the system from previously described PKS systems (e.g., type I, type II or modular). More particularly, the PUFA PKS system described herein contains domains that appear to function during each cycle as well as those which appear to function during only some of the cycles. A key aspect of this functionality may be related to the domains showing homology to the bacterial Fab-A enzymes. For example, the Fab-A enzyme of *E. coli* has been shown to possess two enzymatic activities. It possesses a dehydration activity in which a water molecule (H_2O) is abstracted from a carbon chain containing a hydroxy group, leaving a *trans* double bond in that carbon chain. In addition, it has an isomerase activity in which the *trans* double bond is converted to the *cis* configuration. This isomerization is accomplished in conjunction with a migration of the double bond position to adjacent carbons. In PKS (and FAS) systems, the main carbon chain

is extended in 2 carbon increments. One can therefore predict the number of extension reactions required to produce the PUFA products of these PKS systems. For example, to produce DHA (C22:6, all *cis*) requires 10 extension reactions. Since there are only 6 double bonds in the end product, it means that during some of the reaction cycles, a double bond is retained (as a *cis* isomer), and in others, the double bond is reduced prior to the next extension.

Before the discovery of a PUFA PKS system in marine bacteria (see U.S. Patent No. 6,140,486), PKS systems were not known to possess this combination of iterative and selective enzymatic reactions, and they were not thought of as being able to produce carbon-carbon double bonds in the *cis* configuration. However, the PUFA PKS system described by the present invention has the capacity to introduce *cis* double bonds and the capacity to vary the reaction sequence in the cycle.

The present inventors propose to use these features of the PUFA PKS system to produce a range of bioactive molecules that could not be produced by the previously described (Type II, Type I and modular) PKS systems. These bioactive molecules include, but are not limited to, polyunsaturated fatty acids (PUFAs), antibiotics or other bioactive compounds, many of which will be discussed below. For example, using the knowledge of the PUFA PKS gene structures described herein, any of a number of methods can be used to alter the PUFA PKS genes, or combine portions of these genes with other synthesis systems, including other PKS systems, such that new products are produced. The inherent ability of this particular type of system to do both iterative and selective reactions will enable this system to yield products that would not be found if similar methods were applied to other types of PKS systems.

Much of the structure of the PKS system for PUFA synthesis in the eukaryotic Thraustochytrid, *Schizochytrium* has been described in detail in U.S. Patent No. 6,566,583. Complete sequencing of cDNA and genomic clones in *Schizochytrium* by the present inventors allowed the identification of the full-length genomic sequence of each of OrfA, OrfB and OrfC and the complete identification of the specific domains in these *Schizochytrium* Orfs with homology to those in *Shewanella* (see Fig. 2 and U.S. Patent

Application Serial No. 10/124,800, *supra*). In U.S. Patent Application Serial No. 10/124,800, the inventors also identified a *Thraustochytrium* species as meeting the criteria for having a PUFA PKS system and then demonstrated that this organism was likely to contain genes with homology to *Schizochytrium* PUFA PKS genes by Southern blot analysis. However, the isolation and determination of the structure of such genes and the domain organization of the genes was not described in U.S. Patent Application Serial No. 10/124,800. In the present invention, the inventors have now cloned and sequenced the full-length genomic sequence of homologous open reading frames (Orfs) in this Thraustochytrid of the genus *Thraustochytrium* (specifically, *Thraustochytrium* sp. 23B (ATCC 20892)), and have identified the domains comprising the PUFA PKS system in this *Thraustochytrium*. Therefore, the present invention solves the above-mentioned problem of providing additional PUFA PKS systems that have the flexibility for commercial use. The *Thraustochytrium* PUFA PKS system is described in detail below.

The present invention also solves the above-identified problem for production of commercially valuable lipids enriched in a desired PUFA, such as EPA, by the present inventors' development of genetically modified microorganisms and methods for efficiently producing lipids (triacylglyerols (TAG) as well as membrane-associated phospholipids (PL)) enriched in PUFAs by manipulation of the polyketide synthase-like system that produces PUFAs in eukaryotes, including members of the order Thraustochytriales such as *Schizochytrium* and *Thraustochytrium*. Specifically, and by way of example, the present inventors describe herein a strain of *Schizochytrium* that has previously been optimized for commercial production of oils enriched in PUFA, primarily docosahexaenoic acid (DHA; C22:6 n-3) and docosapentaenoic acid (DPA; C22:5 n-6), and that will now be genetically modified such that EPA (C20:5 n-3) production (or other PUFA production) replaces the DHA production, without sacrificing the oil productivity characteristics of the organism. In addition, the present inventors describe herein the genetic modification of *Schizochytrium* with PUFA PKS genes from *Thraustochytrium* to improve the DHA production by the *Schizochytrium* organism, specifically by altering the ratio of DHA to DPA produced by the microorganism through the modification of the PUFA PKS system. These are only a few

examples of the technology encompassed by the invention, as the concepts of the invention can readily be applied to other production organisms and other desired PUFAs as described in detail below.

In one embodiment, a PUFA PKS system according to the present invention comprises at least the following biologically active domains: (a) at least two enoyl-ACP reductase (ER) domains; (b) at least six acyl carrier protein (ACP) domains; (c) at least two β-ketoacyl-ACP synthase (KS) domains; (d) at least one acyltransferase (AT) domain; (e) at least one β-ketoacyl-ACP reductase (KR) domain; (f) at least two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) at least one chain length factor (CLF) domain; and (h) at least one malonyl-CoA:ACP acyltransferase (MAT) domain. The functions of these domains are generally individually known in the art and will be described in detail below with regard to the PUFA PKS system of the present invention.

In another embodiment, the PUFA PKS system comprises at least the following biologically active domains: (a) at least one enoyl-ACP reductase (ER) domain; (b) multiple acyl carrier protein (ACP) domains (at least from one to four, and preferably at least five, and more preferably at least six, and even more preferably seven, eight, nine, or more than nine); (c) at least two β-ketoacyl-ACP synthase (KS) domains; (d) at least one acyltransferase (AT) domain; (e) at least one β-ketoacyl-ACP reductase (KR) domain; (f) at least two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; (g) at least one chain length factor (CLF) domain; and (h) at least one malonyl-CoA:ACP acyltransferase (MAT) domain. Preferably, such a PUFA PKS system is a non-bacterial PUFA-PKS system.

In one embodiment, a PUFA PKS system of the present invention is a non-bacterial PUFA PKS system. In other words, in one embodiment, the PUFA PKS system of the present invention is isolated from an organism that is not a bacterium, or is a homologue of, or derived from, a PUFA PKS system from an organism that is not a bacterium, such as a eukaryote or an archaebacterium. Eukaryotes are separated from prokaryotes based on the degree of differentiation of the cells, with eukaryotes having more highly differentiated cells and prokaryotes having less differentiated cells. In general, prokaryotes do not possess a nuclear membrane, do not exhibit mitosis during cell division, have only one chromosome,

their cytoplasm contains 70S ribosomes, they do not possess any mitochondria, endoplasmic reticulum, chloroplasts, lysosomes or Golgi apparatus, their flagella (if present) consists of a single fibril. In contrast, eukaryotes have a nuclear membrane, they do exhibit mitosis during cell division, they have many chromosomes, their cytoplasm contains 80S ribosomes, they do possess mitochondria, endoplasmic reticulum, chloroplasts (in algae), lysosomes and Golgi apparatus, and their flagella (if present) consists of many fibrils. In general, bacteria are prokaryotes, while algae, fungi, protist, protozoa and higher plants are eukaryotes.

The PUFA PKS systems of the marine bacteria (e.g., *Shewanella* sp. strain SCRC2738 and *Vibrio marinus*) are not the basis of the present invention, although the present invention does contemplate the use of domains from these bacterial PUFA PKS systems *in conjunction* with domains from the non-bacterial PUFA PKS systems of the present invention. In addition, the present invention does contemplate the isolation and use of PUFA PKS gene sets (and proteins and domains encoded thereby) isolated from other bacteria (e.g. *Shewanella olleyana* and *Shewanella japonica*) that will be particularly suitable for use as sources of PUFA PKS genes for modifying or combining with the non-bacterial PUFA PKS genes described herein to produce hybrid constructs and genetically modified microorganisms and plants. For example, according to the present invention, genetically modified organisms can be produced which incorporate non-bacterial PUFA PKS functional domains with bacterial PUFA PKS functional domains, as well as PKS functional domains or proteins from other PKS systems (type I, type II, modular) or FAS systems. As discussed in more detail below, PUFA PKS genes from two species of *Shewanella*, namely *Shewanella olleyana* or *Shewanella japonica*, are exemplary bacterial genes that are preferred for use in genetically modified microorganisms, plants, and methods of the invention. PUFA PKS systems (genes and the proteins and domains encoded thereby) from such marine bacteria (e.g., *Shewanella olleyana* or *Shewanella japonica*) are encompassed by the present invention as novel PUFA PKS sequences.

According to the present invention, the terms/phrases "Thraustochytrid", "Thraustochytriales microorganism" and "microorganism of the order Thraustochytriales" can be used interchangeably and refer to any members of the order Thraustochytriales, which

includes both the family Thraustochytriaceae and the family Labyrinthulaceae. The terms "Labyrinthulid" and "Labyrinthulaceae" are used herein to specifically refer to members of the family Labyrinthulaceae. To specifically reference Thraustochytrids that are members of the family Thraustochytriaceae, the term "Thraustochytriaceae" is used herein. Thus, for the present invention, members of the Labyrinthulids are considered to be included in the Thraustochytrids.

Developments have resulted in frequent revision of the taxonomy of the Thraustochytrids. Taxonomic theorists generally place Thraustochytrids with the algae or algae-like protists. However, because of taxonomic uncertainty, it would be best for the purposes of the present invention to consider the strains described in the present invention as Thraustochytrids to include the following organisms: Order: Thraustochytriales; Family: Thraustochytriaceae (Genera: *Thraustochytrium*, *Schizochytrium*,, *Japonochytrium*, *Aplanochytrium*, or *Elina*) or Labyrinthulaceae (Genera *Labyrinthula*, *Labyrinthuloides*, or *Labyrinthomyxa*). Also, the following genera are sometimes included in either family Thraustochytriaceae or Labyrinthulaceae: *Althornia, Corallochytrium, Diplophyrys, and Pyrrhosorus*), and for the purposes of this invention are encompassed by reference to a Thraustochytrid or a member of the order Thraustochytriales. It is recognized that at the time of this invention, revision in the taxonomy of Thraustochytrids places the genus *Labyrinthuloides* in the family of Labyrinthulaceae and confirms the placement of the two families Thraustochytriaceae and Labyrinthulaceae within the Stramenopile lineage. It is noted that the Labyrinthulaceae are sometimes commonly called labyrinthulids or labyrinthula, or labyrinthuloides and the Thraustochytriaceae are commonly called thraustochytrids, although, as discussed above, for the purposes of clarity of this invention, reference to Thraustochytrids encompasses any member of the order Thraustochytriales and/or includes members of both Thraustochytriaceae and Labyrinthulaceae. Recent taxonomic changes are summarized below.

Strains of certain unicellular microorganisms disclosed herein are members of the order Thraustochytriales. Thraustochytrids are marine eukaryotes with an evolving taxonomic history. Problems with the taxonomic placement of the Thraustochytrids have

been reviewed by Moss (1986), Bahnweb and Jackle (1986) and Chamberlain and Moss (1988).

For convenience purposes, the Thraustochytrids were first placed by taxonomists with other colorless zoosporic eukaryotes in the Phycomycetes (algae-like fungi). The name Phycomycetes, however, was eventually dropped from taxonomic status, and the Thraustochytrids were retained in the Oomycetes (the biflagellate zoosporic fungi). It was initially assumed that the Oomycetes were related to the heterokont algae, and eventually a wide range of ultrastructural and biochemical studies, summarized by Barr (Barr, 1981, *Biosystems* 14:359-370) supported this assumption. The Oomycetes were in fact accepted by Leedale (Leedale, 1974, *Taxon* 23:261-270) and other phycologists as part of the heterokont algae. However, as a matter of convenience resulting from their heterotrophic nature, the Oomycetes and Thraustochytrids have been largely studied by mycologists (scientists who study fungi) rather than phycologists (scientists who study algae).

From another taxonomic perspective, evolutionary biologists have developed two general schools of thought as to how eukaryotes evolved. One theory proposes an exogenous origin of membrane-bound organelles through a series of endosymbioses (Margulis, 1970, Origin of Eukaryotic Cells. Yale University Press, New Haven); e.g., mitochondria were derived from bacterial endosymbionts, chloroplasts from cyanophytes, and flagella from spirochaetes. The other theory suggests a gradual evolution of the membrane-bound organelles from the non-membrane-bounded systems of the prokaryote ancestor via an autogenous process (Cavalier-Smith, 1975, *Nature* (Lond.) 256:462-468). Both groups of evolutionary biologists however, have removed the Oomycetes and Thraustochytrids from the fungi and place them either with the chromophyte algae in the kingdom Chromophyta (Cavalier-Smith, 1981, *BioSystems* 14:461-481) (this kingdom has been more recently expanded to include other protists and members of this kingdom are now called Stramenopiles) or with all algae in the kingdom Protoctista (Margulis and Sagen, 1985, Biosystems 18:141-147).

With the development of electron microscopy, studies on the ultrastructure of the zoospores of two genera of Thraustochytrids, *Thraustochytrium* and *Schizochytrium*,

(Perkins, 1976, pp. 279-312 in "Recent Advances in Aquatic Mycology" (ed. E.B.G. Jones), John Wiley & Sons, New York; Kazama, 1980, *Can. J. Bot.* 58:2434-2446; Barr, 1981, *Biosystems* 14:359-370) have provided good evidence that the Thraustochytriaceae are only distantly related to the Oomycetes. Additionally, genetic data representing a correspondence analysis (a form of multivariate statistics) of 5-S ribosomal RNA sequences indicate that Thraustochytriales are clearly a unique group of eukaryotes, completely separate from the fungi, and most closely related to the red and brown algae, and to members of the Oomycetes (Mannella, et al., 1987, *Mol. Evol.* 24:228-235). Most taxonomists have agreed to remove the Thraustochytrids from the Oomycetes (Bartnicki-Garcia, 1987, pp. 389-403 in "Evolutionary Biology of the Fungi" (eds. Rayner, A.D.M., Brasier, C.M. & Moore, D.), Cambridge University Press, Cambridge).

In summary, employing the taxonomic system of Cavalier-Smith (Cavalier-Smith, 1981, *BioSystems* 14:461-481, 1983; Cavalier-Smith, 1993, *Microbiol Rev.* 57:953-994), the Thraustochytrids are classified with the chromophyte algae in the kingdom Chromophyta (Stramenopiles). This taxonomic placement has been more recently reaffirmed by Cavalier-Smith et al. using the 18s rRNA signatures of the Heterokonta to demonstrate that Thraustochytrids are chromists not Fungi (Cavalier-Smith et al., 1994, *Phil. Tran. Roy. Soc. London Series BioSciences* 346:387-397). This places the Thraustochytrids in a completely different kingdom from the fungi, which are all placed in the kingdom Eufungi.

Currently, there are 71 distinct groups of eukaryotic organisms (Patterson 1999) and within these groups four major lineages have been identified with some confidence: (1) Alveolates, (2) Stramenopiles, (3) a Land Plant-green algae-Rhodophyte_Glaucophyte ("plant") clade and (4) an Opisthokont clade (Fungi and Animals). Formerly these four major lineages would have been labeled Kingdoms but use of the "kingdom" concept is no longer considered useful by some researchers.

As noted by Armstrong, Stramenopile refers to three-parted tubular hairs, and most members of this lineage have flagella bearing such hairs. Motile cells of the Stramenopiles (unicellular organisms, sperm, zoopores) are asymmetrical having two laterally inserted flagella, one long, bearing three-parted tubular hairs that reverse the thrust of the flagellum,

and one short and smooth. Formerly, when the group was less broad, the Stramenopiles were called Kingdom Chromista or the heterokont (=different flagella) algae because those groups consisted of the Brown Algae or Phaeophytes, along with the yellow-green Algae, Golden-brown Algae, Eustigmatophytes and Diatoms. Subsequently some heterotrophic, fungal-like organisms, the water molds, and labyrinthulids (slime net amoebas), were found to possess similar motile cells, so a group name referring to photosynthetic pigments or algae became inappropriate. Currently, two of the families within the Stramenopile lineage are the Labyrinthulaceae and the Thraustochytriaceae. Historically, there have been numerous classification strategies for these unique microorganisms and they are often classified under the same order (i.e., Thraustochytriales). Relationships of the members in these groups are still developing. Porter and Leander have developed data based on 18S small subunit ribosomal DNA indicating the thraustochytrid-labyrinthulid clade in monophyletic. However, the clade is supported by two branches; the first contains three species of *Thraustochytrium* and *Ulkenia profunda*, and the second includes three species of *Labyrinthula*, two species of *Labyrinthuloides* and *Schizochytrium aggregatum*.

The taxonomic placement of the Thraustochytrids as used in the present invention is therefore summarized below:

Kingdom: Chromophyta (Stramenopiles)
Phylum: Heterokonta
Order: Thraustochytriales (Thraustochytrids)
Family: Thraustochytriaceae or Labyrinthulaceae
Genera: *Thraustochytrium*, *Schizochytrium*, *Japonochytrium*, *Aplanochytrium, Elina*, *Labyrinthula, Labyrinthuloides,* or *Labyrinthulomyxa*

Some early taxonomists separated a few original members of the genus *Thraustochytrium* (those with an amoeboid life stage) into a separate genus called *Ulkenia*. However it is now known that most, if not all, Thraustochytrids (including *Thraustochytrium* and *Schizochytrium*), exhibit amoeboid stages and as such, *Ulkenia* is not considered by some to be a valid genus. As used herein, the genus *Thraustochytrium* will include *Ulkenia*.

Despite the uncertainty of taxonomic placement within higher classifications of Phylum and Kingdom, the Thraustochytrids remain a distinctive and characteristic grouping whose members remain classifiable within the order Thraustochytriales.

Schizochytrium is a Thraustochytrid marine microorganism that accumulates large quantities of triacylglycerols rich in DHA and docosapentaenoic acid (DPA; 22:5 ω-6); e.g., 30% DHA + DPA by dry weight (Barclay et al., *J. Appl. Phycol.* **6**, 123 (1994)). In eukaryotes that synthesize 20- and 22-carbon PUFAs by an elongation/desaturation pathway, the pools of 18-, 20- and 22-carbon intermediates are relatively large so that *in vivo* labeling experiments using [^{14}C]-acetate reveal clear precursor-product kinetics for the predicted intermediates (Gellerman et al., *Biochim. Biophys. Acta* 573:23 (1979)). Furthermore, radiolabeled intermediates provided exogenously to such organisms are converted to the final PUFA products. The present inventors have shown that [1-^{14}C]-acetate was rapidly taken up by *Schizochytrium* cells and incorporated into fatty acids, but at the shortest labeling time (1 min), DHA contained 31% of the label recovered in fatty acids, and this percentage remained essentially unchanged during the 10-15 min of [^{14}C]-acetate incorporation and the subsequent 24 hours of culture growth. Similarly, DPA represented 10% of the label throughout the experiment. There is no evidence for a precursor-product relationship between 16- or 18-carbon fatty acids and the 22-carbon polyunsaturated fatty acids. These results are consistent with rapid synthesis of DHA from [^{14}C]-acetate involving very small (possibly enzyme-bound) pools of intermediates. A cell-free homogenate derived from *Schizochytrium* cultures incorporated [1-^{14}C]-malonyl-CoA into DHA, DPA, and saturated fatty acids. The same biosynthetic activities were retained by a 100,000xg supernatant fraction but were not present in the membrane pellet. Thus, DHA and DPA synthesis in *Schizochytrium* does not involve membrane-bound desaturases or fatty acid elongation enzymes like those described for other eukaryotes (Parker-Barnes et al., 2000, *supra*; Shanklin et al., 1998, *supra*). These fractionation data contrast with those obtained from the *Shewanella* enzymes (See Metz et al., 2001, *supra*) and may indicate use of a different (soluble) acyl acceptor molecule, such as CoA, by the *Schizochytrium* enzyme. It is expected that *Thraustochytrium* will have a similar biochemistry.

In U.S. Patent No. 6,566,583, a cDNA library from *Schizochytrium* was constructed and approximately 8500 random clones (ESTs) were sequenced. Sequences that exhibited homology to 8 of the 11 domains of the *Shewanella* PKS genes shown in Fig. 2 were all identified at frequencies of 0.2-0.5%. In U.S. Patent No. 6,566,583, several cDNA clones from *Schizochytrium* showing homology to the *Shewanella* PKS genes were sequenced, and various clones were assembled into nucleic acid sequences representing two partial open reading frames and one complete open reading frame.

Further sequencing of cDNA and genomic clones by the present inventors allowed the identification of the full-length genomic sequence of each of OrfA, OrfB and OrfC in *Schizochytrium* and the complete identification of the domains in *Schizochytrium* with homology to those in *Shewanella* (see Fig. 2). These genes are described in detail in U.S. Patent Application Serial No. 10/124,800, *supra* and are described in some detail below.

The present inventors have now identified, cloned, and sequenced the full-length genomic sequence of homologous Orfs in a Thraustochytrid of the genus *Thraustochytrium* (specifically, *Thraustochytrium* sp. 23B (ATCC 20892)) and have identified the domains comprising the PUFA PKS system in this *Thraustochytrium*.

Based on the comparison of the domains of the PUFA PKS system of *Schizochytrium* with the domains of the PUFA PKS system of *Shewanella*, clearly, the *Schizochytrium* genome encodes proteins that are highly similar to the proteins in *Shewanella* that are capable of catalyzing EPA synthesis. The proteins in *Schizochytrium* constitute a PUFA PKS system that catalyzes DHA and DPA synthesis. Simple modification of the reaction scheme identified for *Shewanella* will allow for DHA synthesis in *Schizochytrium*. The homology between the prokaryotic *Shewanella* and eukaryotic *Schizochytrium* genes suggests that the PUFA PKS has undergone lateral gene transfer.

A similar comparison can be made for *Thraustochytrium*. In all cases, comparison of the *Thraustochytrium* 23B (*Th.* 23B) PUFA PKS proteins or domains to other known sequences revealed that the closest match was one of the *Schizochytrium* PUFA PKS proteins (OrfA, B or C, or a domain therefrom) as described in U.S. Patent Application Serial No. 10/124,800, *supra*. The next closest matches in all cases were to one of the PUFA PKS

proteins from marine bacteria (*Shewanella* SCRC-2738, *Shewanella oneidensis*, *Photobacter profundum* and *Moritella marina*) or from a related system found in nitrogen fixing cyanobacteria (e.g., *Nostoc punctiforme* and *Nostoc* sp. PCC 7120). The products of the cyanobacterial enzyme systems lack double bonds and the proteins lack domains related to the DH domains implicated in *cis* double bond formation (i.e., the FabA related DH domains).

According to the present invention, the phrase "open reading frame" is denoted by the abbreviation "Orf". It is noted that the protein encoded by an open reading frame can also be denoted in all upper case letters as "ORF" and a nucleic acid sequence for an open reading frame can also be denoted in all lower case letters as "orf", but for the sake of consistency, the spelling "Orf" is preferentially used herein to describe either the nucleic acid sequence or the protein encoded thereby. It will be obvious from the context of the usage of the term whether a protein or nucleic acid sequence is referenced.

***Schizochytrium* PUFA PKS**

Fig. 1 is a graphical representation of the three open reading frames from the *Schizochytrium* PUFA PKS system, and includes the domain structure of this PUFA PKS system. As described in detail in U.S. Patent Application Serial No. 10/124,800, the domain structure of each open reading frame is as follows:

Open Reading Frame A (OrfA):

The complete nucleotide sequence for OrfA is represented herein as SEQ ID NO:1. OrfA is a 8730 nucleotide sequence (not including the stop codon) which encodes a 2910 amino acid sequence, represented herein as SEQ ID NO:2. Within OrfA are twelve domains: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one malonyl-CoA:ACP acyltransferase (MAT) domain; (c) nine acyl carrier protein (ACP) domains; and (d) one β-ketoacyl-ACP reductase (KR) domain. The nucleotide sequence for OrfA has been deposited with GenBank as Accession No. AF378327 (amino acid sequence Accession No. AAK728879).

The first domain in *Schizochytrium* OrfA is a β-ketoacyl-ACP synthase (KS) domain, also referred to herein as OrfA-KS. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1 and 40 of SEQ ID NO:1 (OrfA)

to an ending point of between about positions 1428 and 1500 of SEQ ID NO:1. The nucleotide sequence containing the sequence encoding the OrfA-KS domain is represented herein as SEQ ID NO:7 (positions 1-1500 of SEQ ID NO:1). The amino acid sequence containing the KS domain spans from a starting point of between about positions 1 and 14 of SEQ ID NO:2 (OrfA) to an ending point of between about positions 476 and 500 of SEQ ID NO:2. The amino acid sequence containing the OrfA-KS domain is represented herein as SEQ ID NO:8 (positions 1-500 of SEQ ID NO:2). It is noted that the OrfA-KS domain contains an active site motif: DXAC* (*acyl binding site C_{215}).

According to the present invention, a domain or protein having β-ketoacyl-ACP synthase (KS) biological activity (function) is characterized as the enzyme that carries out the initial step of the FAS (and PKS) elongation reaction cycle. The term "β-ketoacyl-ACP synthase" can be used interchangeably with the terms "3-keto acyl-ACP synthase", "β-keto acyl-ACP synthase", and "keto-acyl ACP synthase", and similar derivatives. The acyl group destined for elongation is linked to a cysteine residue at the active site of the enzyme by a thioester bond. In the multi-step reaction, the acyl-enzyme undergoes condensation with malonyl-ACP to form -ketoacyl-ACP, CO_2 and free enzyme. The KS plays a key role in the elongation cycle and in many systems has been shown to possess greater substrate specificity than other enzymes of the reaction cycle. For example, *E. coli* has three distinct KS enzymes - each with its own particular role in the physiology of the organism (Magnuson et al., *Microbiol. Rev.* **57**, 522 (1993)). The two KS domains of the PUFA-PKS systems could have distinct roles in the PUFA biosynthetic reaction sequence.

As a class of enzymes, KS's have been well characterized. The sequences of many verified KS genes are known, the active site motifs have been identified and the crystal structures of several have been determined. Proteins (or domains of proteins) can be readily identified as belonging to the KS family of enzymes by homology to known KS sequences.

The second domain in OrfA is a malonyl-CoA:ACP acyltransferase (MAT) domain, also referred to herein as OrfA-MAT. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1723 and 1798 of SEQ ID NO:1 (OrfA) to an ending point of between about positions 2805 and 3000 of SEQ ID

NO:1. The nucleotide sequence containing the sequence encoding the OrfA-MAT domain is represented herein as SEQ ID NO:9 (positions 1723-3000 of SEQ ID NO:1). The amino acid sequence containing the MAT domain spans from a starting point of between about positions 575 and 600 of SEQ ID NO:2 (OrfA) to an ending point of between about positions 935 and 1000 of SEQ ID NO:2. The amino acid sequence containing the OrfA-MAT domain is represented herein as SEQ ID NO:10 (positions 575-1000 of SEQ ID NO:2). It is noted that the OrfA-MAT domain contains an active site motif: GHS*XG (*acyl binding site S_{706}), represented herein as SEQ ID NO:11.

According to the present invention, a domain or protein having malonyl-CoA:ACP acyltransferase (MAT) biological activity (function) is characterized as one that transfers the malonyl moiety from malonyl-CoA to ACP. The term "malonyl-CoA:ACP acyltransferase" can be used interchangeably with "malonyl acyltransferase" and similar derivatives. In addition to the active site motif (GxSxG), these enzymes possess an extended motif (R and Q amino acids in key positions) that identifies them as MAT enzymes (in contrast to the AT domain of Schizochytrium Orf B). In some PKS systems (but not the PUFA PKS domain) MAT domains will preferentially load methyl- or ethyl- malonate on to the ACP group (from the corresponding CoA ester), thereby introducing branches into the linear carbon chain. MAT domains can be recognized by their homology to known MAT sequences and by their extended motif structure.

Domains 3-11 of OrfA are nine tandem acyl carrier protein (ACP) domains, also referred to herein as OrfA-ACP (the first domain in the sequence is OrfA-ACP1, the second domain is OrfA-ACP2, the third domain is OrfA-ACP3, etc.). The first ACP domain, OrfA-ACP1, is contained within the nucleotide sequence spanning from about position 3343 to about position 3600 of SEQ ID NO:1 (OrfA). The nucleotide sequence containing the sequence encoding the OrfA-ACP1 domain is represented herein as SEQ ID NO:12 (positions 3343-3600 of SEQ ID NO:1). The amino acid sequence containing the first ACP domain spans from about position 1115 to about position 1200 of SEQ ID NO:2. The amino acid sequence containing the OrfA-ACP1 domain is represented herein as SEQ ID NO:13 (positions 1115-1200 of SEQ ID NO:2). It is noted that the OrfA-ACP1 domain contains an

active site motif: LGIDS* (*pantetheine binding motif S_{1157}), represented herein by SEQ ID NO:14.

The nucleotide and amino acid sequences of all nine ACP domains are highly conserved and therefore, the sequence for each domain is not represented herein by an individual sequence identifier. However, based on the information disclosed herein, one of skill in the art can readily determine the sequence containing each of the other eight ACP domains (see discussion below).

All nine ACP domains together span a region of OrfA of from about position 3283 to about position 6288 of SEQ ID NO:1, which corresponds to amino acid positions of from about 1095 to about 2096 of SEQ ID NO:2. The nucleotide sequence for the entire ACP region containing all nine domains is represented herein as SEQ ID NO:16. The region represented by SEQ ID NO:16 includes the linker segments between individual ACP domains. The repeat interval for the nine domains is approximately every 330 nucleotides of SEQ ID NO:16 (the actual number of amino acids measured between adjacent active site serines ranges from 104 to 116 amino acids). Each of the nine ACP domains contains a pantetheine binding motif LGIDS* (represented herein by SEQ ID NO:14), wherein S* is the pantetheine binding site serine (S). The pantetheine binding site serine (S) is located near the center of each ACP domain sequence. At each end of the ACP domain region and between each ACP domain is a region that is highly enriched for proline (P) and alanine (A), which is believed to be a linker region. For example, between ACP domains 1 and 2 is the sequence: APAPVKAAAPAAPVASAPAPA, represented herein as SEQ ID NO:15. The locations of the active site serine residues (i.e., the pantetheine binding site) for each of the nine ACP domains, with respect to the amino acid sequence of SEQ ID NO:2, are as follows: ACP1 = S_{1157}; ACP2 = S_{1266}; ACP3 = S_{1377}; ACP4 = S_{1488}; ACP5 = S_{1604}; ACP6 = S_{1715}; ACP7 = S_{1819}; ACP8 = S_{1930}; and ACP9 = S_{2034}. Given that the average size of an ACP domain is about 85 amino acids, excluding the linker, and about 110 amino acids including the linker, with the active site serine being approximately in the center of the domain, one of skill in the art can readily determine the positions of each of the nine ACP domains in OrfA.

According to the present invention, a domain or protein having acyl carrier protein (ACP) biological activity (function) is characterized as being small polypeptides (typically, 80 to 100 amino acids long), that function as carriers for growing fatty acyl chains via a thioester linkage to a covalently bound co-factor of the protein. They occur as separate units or as domains within larger proteins. ACPs are converted from inactive apo-forms to functional holo-forms by transfer of the phosphopantetheinyl moeity of CoA to a highly conserved serine residue of the ACP. Acyl groups are attached to ACP by a thioester linkage at the free terminus of the phosphopantetheinyl moiety. ACPs can be identified by labeling with radioactive pantetheine and by sequence homology to known ACPs. The presence of variations of the above mentioned motif (LGIDS*) is also a signature of an ACP.

Domain 12 in OrfA is a β-ketoacyl-ACP reductase (KR) domain, also referred to herein as OrfA-KR. This domain is contained within the nucleotide sequence spanning from a starting point of about position 6598 of SEQ ID NO:1 to an ending point of about position 8730 of SEQ ID NO:1. The nucleotide sequence containing the sequence encoding the OrfA-KR domain is represented herein as SEQ ID NO:17 (positions 6598-8730 of SEQ ID NO:1). The amino acid sequence containing the KR domain spans from a starting point of about position 2200 of SEQ ID NO:2 (OrfA) to an ending point of about position 2910 of SEQ ID NO:2. The amino acid sequence containing the OrfA-KR domain is represented herein as SEQ ID NO:18 (positions 2200-2910 of SEQ ID NO:2). Within the KR domain is a core region with homology to short chain aldehyde-dehydrogenases (KR is a member of this family). This core region spans from about position 7198 to about position 7500 of SEQ ID NO:1, which corresponds to amino acid positions 2400-2500 of SEQ ID NO:2.

According to the present invention, a domain or protein having β-ketoacyl-ACP reductase (KR) activity is characterized as one that catalyzes the pyridine-nucleotide-dependent reduction of 3-ketoacyl forms of ACP. The term "β-ketoacyl-ACP reductase" can be used interchangeably with the terms "ketoreductase", "3-ketoacyl-ACP reductase", "keto-acyl ACP reductase" and similar derivatives of the term. It is the first reductive step in the *de novo* fatty acid biosynthesis elongation cycle and a reaction often performed in polyketide biosynthesis. Significant sequence similarity is observed with one family of enoyl-ACP

reductases (ER), the other reductase of FAS (but not the ER family present in the PUFA PKS system), and the short-chain alcohol dehydrogenase family. Pfam analysis of the PUFA PKS region indicated above reveals the homology to the short-chain alcohol dehydrogenase family in the core region. Blast analysis of the same region reveals matches in the core area to known KR enzymes as well as an extended region of homology to domains from the other characterized PUFA PKS systems.

Open Reading Frame B (OrfB):

The complete nucleotide sequence for OrfB is represented herein as SEQ ID NO:3. OrfB is a 6177 nucleotide sequence (not including the stop codon) which encodes a 2059 amino acid sequence, represented herein as SEQ ID NO:4. Within OrfB are four domains: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one chain length factor (CLF) domain; (c) one acyltransferase (AT) domain; and, (d) one enoyl-ACP reductase (ER) domain. The nucleotide sequence for OrfB has been deposited with GenBank as Accession No. AF378328 (amino acid sequence Accession No. AAK728880).

The first domain in OrfB is a β-ketoacyl-ACP synthase (KS) domain, also referred to herein as OrfB-KS. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1 and 43 of SEQ ID NO:3 (OrfB) to an ending point of between about positions 1332 and 1350 of SEQ ID NO:3. The nucleotide sequence containing the sequence encoding the OrfB-KS domain is represented herein as SEQ ID NO:19 (positions 1-1350 of SEQ ID NO:3). The amino acid sequence containing the KS domain spans from a starting point of between about positions 1 and 15 of SEQ ID NO:4 (OrfB) to an ending point of between about positions 444 and 450 of SEQ ID NO:4. The amino acid sequence containing the OrfB-KS domain is represented herein as SEQ ID NO:20 (positions 1-450 of SEQ ID NO:4). It is noted that the OrfB-KS domain contains an active site motif: DXAC* (*acyl binding site C_{196}). KS biological activity and methods of identifying proteins or domains having such activity is described above.

The second domain in OrfB is a chain length factor (CLF) domain, also referred to herein as OrfB-CLF. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1378 and 1402 of SEQ ID NO:3 (OrfB) to

an ending point of between about positions 2682 and 2700 of SEQ ID NO:3. The nucleotide sequence containing the sequence encoding the OrfB-CLF domain is represented herein as SEQ ID NO:21 (positions 1378-2700 of SEQ ID NO:3). The amino acid sequence containing the CLF domain spans from a starting point of between about positions 460 and 468 of SEQ ID NO:4 (OrfB) to an ending point of between about positions 894 and 900 of SEQ ID NO:4. The amino acid sequence containing the OrfB-CLF domain is represented herein as SEQ ID NO:22 (positions 460-900 of SEQ ID NO:4). It is noted that the OrfB-CLF domain contains a KS active site motif without the acyl-binding cysteine.

According to the present invention, a domain or protein is referred to as a chain length factor (CLF) based on the following rationale. The CLF was originally described as characteristic of Type II (dissociated enzymes) PKS systems and was hypothesized to play a role in determining the number of elongation cycles, and hence the chain length, of the end product. CLF amino acid sequences show homology to KS domains (and are thought to form heterodimers with a KS protein), but they lack the active site cysteine. CLF's role in PKS systems is currently controversial. New evidence (C. Bisang et al., *Nature* **401**, 502 (1999)) suggests a role in priming (providing the initial acyl group to be elongated) the PKS systems. In this role the CLF domain is thought to decarboxylate malonate (as malonyl-ACP), thus forming an acetate group that can be transferred to the KS active site. This acetate therefore acts as the 'priming' molecule that can undergo the initial elongation (condensation) reaction. Homologues of the Type II CLF have been identified as 'loading' domains in some modular PKS systems. A domain with the sequence features of the CLF is found in all currently identified PUFA PKS systems and in each case is found as part of a multidomain protein.

The third domain in OrfB is an AT domain, also referred to herein as OrfB-AT. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 2701 and 3598 of SEQ ID NO:3 (OrfB) to an ending point of between about positions 3975 and 4200 of SEQ ID NO:3. The nucleotide sequence containing the sequence encoding the OrfB-AT domain is represented herein as SEQ ID NO:23 (positions 2701-4200 of SEQ ID NO:3). The amino acid sequence containing the AT domain spans from a starting point of between about positions 901 and 1200 of SEQ ID

NO:4 (OrfB) to an ending point of between about positions 1325 and 1400 of SEQ ID NO:4. The amino acid sequence containing the OrfB-AT domain is represented herein as SEQ ID NO:24 (positions 901-1400 of SEQ ID NO:4). It is noted that the OrfB-AT domain contains an active site motif of GxS*xG (*acyl binding site S_{1140}) that is characteristic of acyltransferse (AT) proteins.

An "acyltransferase" or "AT" refers to a general class of enzymes that can carry out a number of distinct acyl transfer reactions. The term "acyltransferase" can be used interchangeably with the term "acyl transferase". The *Schizochytrium* domain shows good homology to a domain present in all of the other PUFA PKS systems currently examined and very weak homology to some acyltransferases whose specific functions have been identified (e.g. to malonyl-CoA:ACP acyltransferase, MAT). In spite of the weak homology to MAT, this AT domain is not believed to function as a MAT because it does not possess an extended motif structure characteristic of such enzymes (see MAT domain description, above). For the purposes of this disclosure, the functions of the AT domain in a PUFA PKS system include, but are not limited to: transfer of the fatty acyl group from the OrfA ACP domain(s) to water (i.e. a thioesterase – releasing the fatty acyl group as a free fatty acid), transfer of a fatty acyl group to an acceptor such as CoA, transfer of the acyl group among the various ACP domains, or transfer of the fatty acyl group to a lipophilic acceptor molecule (e.g. to lysophosphadic acid).

The fourth domain in OrfB is an ER domain, also referred to herein as OrfB-ER. This domain is contained within the nucleotide sequence spanning from a starting point of about position 4648 of SEQ ID NO:3 (OrfB) to an ending point of about position 6177 of SEQ ID NO:3. The nucleotide sequence containing the sequence encoding the OrfB-ER domain is represented herein as SEQ ID NO:25 (positions 4648-6177 of SEQ ID NO:3). The amino acid sequence containing the ER domain spans from a starting point of about position 1550 of SEQ ID NO:4 (OrfB) to an ending point of about position 2059 of SEQ ID NO:4. The amino acid sequence containing the OrfB-ER domain is represented herein as SEQ ID NO:26 (positions 1550-2059 of SEQ ID NO:4).

According to the present invention, this domain has enoyl-ACP reductase (ER) biological activity. According to the present invention, the term "enoyl-ACP reductase" can be used interchangeably with "enoyl reductase", "enoyl ACP-reductase" and "enoyl acyl-ACP reductase". The ER enzyme reduces the *trans*-double bond (introduced by the DH activity) in the fatty acyl-ACP, resulting in fully saturating those carbons. The ER domain in the PUFA-PKS shows homology to a newly characterized family of ER enzymes (Heath et al., *Nature* **406**, 145 (2000)). Heath and Rock identified this new class of ER enzymes by cloning a gene of interest from *Streptococcus pneumoniae*, purifying a protein expressed from that gene, and showing that it had ER activity in an *in vitro* assay. The sequence of the Schizochytrium ER domain of OrfB shows homology to the *S. pneumoniae* ER protein. All of the PUFA PKS systems currently examined contain at least one domain with very high sequence homology to the *Schizochytrium* ER domain. The *Schizochytrium* PUFA PKS system contains two ER domains (one on OrfB and one on OrfC).

Open Reading Frame C (OrfC):

The complete nucleotide sequence for OrfC is represented herein as SEQ ID NO:5. OrfC is a 4509 nucleotide sequence (not including the stop codon) which encodes a 1503 amino acid sequence, represented herein as SEQ ID NO:6. Within OrfC are three domains: (a) two FabA-like β-hydroxyacyl-ACP dehydrase (DH) domains; and (b) one enoyl-ACP reductase (ER) domain. The nucleotide sequence for OrfC has been deposited with GenBank as Accession No. AF378329 (amino acid sequence Accession No. AAK728881).

The first domain in OrfC is a DH domain, also referred to herein as OrfC-DH1. This is one of two DH domains in OrfC, and therefore is designated DH1. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1 and 778 of SEQ ID NO:5 (OrfC) to an ending point of between about positions 1233 and 1350 of SEQ ID NO:5. The nucleotide sequence containing the sequence encoding the OrfC-DH1 domain is represented herein as SEQ ID NO:27 (positions 1-1350 of SEQ ID NO:5). The amino acid sequence containing the DH1 domain spans from a starting point of between about positions 1 and 260 of SEQ ID NO:6 (OrfC) to an ending point of between about positions 411 and 450 of SEQ ID NO:6. The amino acid sequence containing the

OrfC-DH1 domain is represented herein as SEQ ID NO:28 (positions 1-450 of SEQ ID NO:6).

According to the present invention, this domain has FabA-like β-hydroxyacyl-ACP dehydrase (DH) biological activity. The term "FabA-like β-hydroxyacyl-ACP dehydrase" can be used interchangeably with the terms "FabA-like β-hydroxy acyl-ACP dehydrase", "β-hydroxyacyl-ACP dehydrase", "dehydrase" and similar derivatives. The characteristics of both the DH domains (see below for DH 2) in the PUFA PKS systems have been described in the preceding sections. This class of enzyme removes HOH from a β-ketoacyl-ACP and leaves a *trans* double bond in the carbon chain. The DH domains of the PUFA PKS systems show homology to bacterial DH enzymes associated with their FAS systems (rather than to the DH domains of other PKS systems). A subset of bacterial DH's, the FabA-like DH's, possesses *cis-trans* isomerase activity (Heath et al., *J. Biol. Chem.*, 271, 27795 (1996)). It is the homologies to the FabA-like DH's that indicate that one or both of the DH domains is responsible for insertion of the cis double bonds in the PUFA PKS products.

The second domain in OrfC is a DH domain, also referred to herein as OrfC-DH2. This is the second of two DH domains in OrfC, and therefore is designated DH2. This domain is contained within the nucleotide sequence spanning from a starting point of between about positions 1351 and 2437 of SEQ ID NO:5 (OrfC) to an ending point of between about positions 2607 and 2850 of SEQ ID NO:5. The nucleotide sequence containing the sequence encoding the OrfC-DH2 domain is represented herein as SEQ ID NO:29 (positions 1351-2850 of SEQ ID NO:5). The amino acid sequence containing the DH2 domain spans from a starting point of between about positions 451 and 813 of SEQ ID NO:6 (OrfC) to an ending point of between about positions 869 and 950 of SEQ ID NO:6. The amino acid sequence containing the OrfC-DH2 domain is represented herein as SEQ ID NO:30 (positions 451-950 of SEQ ID NO:6). DH biological activity has been described above.

The third domain in OrfC is an ER domain, also referred to herein as OrfC-ER. This domain is contained within the nucleotide sequence spanning from a starting point of about position 2998 of SEQ ID NO:5 (OrfC) to an ending point of about position 4509 of SEQ ID

NO:5. The nucleotide sequence containing the sequence encoding the OrfC-ER domain is represented herein as SEQ ID NO:31 (positions 2998-4509 of SEQ ID NO:5). The amino acid sequence containing the ER domain spans from a starting point of about position 1000 of SEQ ID NO:6 (OrfC) to an ending point of about position 1502 of SEQ ID NO:6. The amino acid sequence containing the OrfC-ER domain is represented herein as SEQ ID NO:32 (positions 1000-1502 of SEQ ID NO:6). ER biological activity has been described above.

***Thraustochytrium* 23B PUFA PKS**

Th. 23B *Open Reading Frame A (OrfA)*:

The complete nucleotide sequence for *Th.* 23B OrfA is represented herein as SEQ ID NO:38. SEQ ID NO:38 encodes the following domains in *Th.* 23B OrfA: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one malonyl-CoA:ACP acyltransferase (MAT) domain; (c) eight acyl carrier protein (ACP) domains; and (d) one β-ketoacyl-ACP reductase (KR) domain. This domain organization is the same as is present in *Schizochytrium* Orf A (SEQ ID NO:1) with the exception that the *Th.* 23B Orf A has 8 adjacent ACP domains, while *Schizochytrium* Orf A has 9 adjacent ACP domains. *Th.* 23B OrfA is a 8433 nucleotide sequence (not including the stop codon) which encodes a 2811 amino acid sequence, represented herein as SEQ ID NO:39. The *Th.* 23B OrfA amino acid sequence (SEQ ID NO:39) was compared with known sequences in a standard BLAST search (BLAST parameters: Blastp, low complexity filter Off, program - BLOSUM62,Gap cost - Existence: 11, Extension 1; (BLAST described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety))). At the amino acid level, the sequences with the greatest degree of homology to *Th.* 23B OrfA was *Schizochytrium* Orf A (gb AAK72879.1) (SEQ ID NO:2). The alignment extends over the entire query but is broken into 2 pieces (due to the difference in numbers of ACP repeats). SEQ ID NO:39 first aligns at positions 6 through 1985 (including 8 ACP domains) with SEQ ID NO:2 and shows a sequence identity to SEQ ID NO:2 of 54% over 2017 amino acids. SEQ ID NO:39 also aligns at positions 980 through 2811 with SEQ ID NO:2 and shows a sequence identity to

SEQ ID NO:2 of 43% over 1861 amino acids. In this second alignment, the match is evident for the *Th.* 23B 8X ACPs in the regions of the conserved pantetheine attachment site motif, but is very poor over the 1st *Schizochytrium* ACP domain (i.e., there is not a 9th ACP domain in the *Th.* 23B query sequence, but the Blastp output under theses conditions attempts to align them anyway). SEQ ID NO:39 shows the next closest identity with sequences from *Shewanella oneidensis* (Accession No. NP_717214) and *Photobacter profundum* (Accession No. AAL01060).

The first domain in *Th.* 23B OrfA is a KS domain, also referred to herein as *Th.* 23B OrfA-KS. KS domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from about position 1 to about position 1500 of SEQ ID NO:38, represented herein as SEQ ID NO:40. The amino acid sequence containing the *Th.* 23B KS domain is a region of SEQ ID NO:39 spanning from about position 1 to about position 500 of SEQ ID NO:39, represented herein as SEQ ID NO:41. This region of SEQ ID NO:39 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from position 1 to about position 450 of SEQ ID NO:39 (also positions 1 to about 450 of SEQ ID NO:41). It is noted that the *Th.* 23B OrfA-KS domain contains an active site motif: DXAC* (*acyl binding site C_{207}). Also, a characteristic motif at the end of the *Th.* 23B KS region, GFGG, is present in positions 453-456 of SEQ ID NO:39 (also positions 453-456 of SEQ ID NO:41). The amino acid sequence spanning positions 1-500 of SEQ ID NO:39 is about 79% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 496 amino acids. The amino acid sequence spanning positions 1-450 of SEQ ID NO:39 is about 81% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 446 amino acids.

The second domain in *Th.* 23B OrfA is a MAT domain, also referred to herein as *Th.* 23B OrfA-MAT. MAT domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 1503 and about position 3000 of SEQ ID NO:38, represented herein as SEQ ID NO:42. The amino acid sequence containing the *Th.* 23B MAT domain is a region of SEQ ID NO:39 spanning from about position 501 to about position 1000, represented herein by SEQ ID NO:43. This region of SEQ ID NO:39 has a Pfam match to FabD (malonyl-CoA:ACP acyltransferase)

spanning from about position 580 to about position 900 of SEQ ID NO:39 (positions 80-400 of SEQ ID NO:43). It is noted that the *Th.* 23B OrfA-MAT domain contains an active site motif: GHS*XG (*acyl binding site S_{697}), represented by positions 695-699 of SEQ ID NO:39. The amino acid sequence spanning positions 501-1000 of SEQ ID NO:39 is about 46% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 481 amino acids. The amino acid sequence spanning positions 580-900 of SEQ ID NO:39 is about 50% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 333 amino acids.

Domains 3-10 of *Th.* 23B OrfA are eight tandem ACP domains, also referred to herein as *Th.* 23B OrfA-ACP (the first domain in the sequence is OrfA-ACP1, the second domain is OrfA-ACP2, the third domain is OrfA-ACP3, etc.). The function of ACP domains has been described in detail above. The first *Th.* 23B ACP domain, *Th.* 23B OrfA-ACP1, is contained within the nucleotide sequence spanning from about position 3205 to about position 3555 of SEQ ID NO:38 (OrfA), represented herein as SEQ ID NO:44. The amino acid sequence containing the first *Th.* 23B ACP domain is a region of SEQ ID NO:39 spanning from about position 1069 to about position 1185 of SEQ ID NO:39, represented herein by SEQ ID NO:45. The amino acid sequence spanning positions 1069-1185 of SEQ ID NO:39 is about 65% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 85 amino acids. *Th.* 23B OrfA-ACP1 has a similar identity to any one of the nine ACP domains in *Schizochytrium* OrfA.

The eight ACP domains in *Th.* 23B OrfA are adjacent to one another and can be identified by the presence of the phosphopantetheine binding site motif, LGXDS* (represented by SEQ ID NO:46), wherein the S* is the phosphopantetheine attachment site. The amino acid position of each of the eight S* sites, with reference to SEQ ID NO:39, are 1128 (ACP1), 1244 (ACP2), 1360 (ACP3), 1476 (ACP4), 1592 (ACP5), 1708 (ACP6), 1824 (ACP7) and 1940 (ACP8). The nucleotide and amino acid sequences of all eight *Th.* 23B ACP domains are highly conserved and therefore, the sequence for each domain is not represented herein by an individual sequence identifier. However, based on the information disclosed herein, one of skill in the art can readily determine the sequence containing each of the other seven ACP domains in SEQ ID NO:38 and SEQ ID NO:39.

All eight *Th.* 23B ACP domains together span a region of *Th.* 23B OrfA of from about position 3205 to about postion 5994 of SEQ ID NO:38, which corresponds to amino acid positions of from about 1069 to about 1998 of SEQ ID NO:39. The nucleotide sequence for the entire ACP region containing all eight domains is represented herein as SEQ ID NO:47. SEQ ID NO:47 encodes an amino acid sequence represented herein by SEQ ID NO:48. SEQ ID NO:48 includes the linker segments between individual ACP domains. The repeat interval for the eight domains is approximately every 116 amino acids of SEQ ID NO:48, and each domain can be considered to consist of about 116 amino acids centered on the active site motif (described above). It is noted that the linker regions between the nine adjacent ACP domains in OrfA in *Schizochytrium* are highly enriched in proline and alanine residues, while the linker regions between the eight adjacent ACP domains in OrfA of *Thraustochytrium* are highly enriched in serine residues (and not proline or alanine residues).

The last domain in *Th.* 23B OrfA is a KR domain, also referred to herein as *Th.* 23B OrfA-KR. KR domain function has been discussed in detail above. This domain is contained within the nucleotide sequence spanning from between about position 6001 to about position 8433 of SEQ ID NO:38, represented herein by SEQ ID NO:49. The amino acid sequence containing the *Th.* 23B KR domain is a region of SEQ ID NO:39 spanning from about position 2001 to about position 2811 of SEQ ID NO:39, represented herein by SEQ ID NO:50. This region of SEQ ID NO:39 has a Pfam match to FabG (β-ketoacyl-ACP reductase) spanning from about position 2300 to about 2550 of SEQ ID NO:39 (positions 300-550 of SEQ ID NO:50). The amino acid sequence spanning positions 2001-2811 of SEQ ID NO:39 is about 40% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 831 amino acids. The amino acid sequence spanning positions 2300-2550 of SEQ ID NO:39 is about 51% identical to *Schizochytrium* OrfA (SEQ ID NO:2) over 235 amino acids.

Th. 23B *Open Reading Frame B (OrfB)*:

The complete nucleotide sequence for *Th.* 23B OrfB is represented herein as SEQ ID NO:51. SEQ ID NO:51 encodes the following domains in *Th.* 23B OrfB: (a) one β-ketoacyl-ACP synthase (KS) domain; (b) one chain length factor (CLF) domain; (c) one acyltransferase (AT) domain; and, (d) one enoyl-ACP reductase (ER) domain. This domain

organization is the same as in *Schizochytrium* Orf B (SEQ ID NO:3) with the exception that the linker region between the AT and ER domains of the *Schizochytrium* protein is longer than that of *Th.* 23B by about 50-60 amino acids. Also, this linker region in *Schizochytrium* has a specific area that is highly enriched in serine residues (it contains 15 adjacent serine residues, in addition to other serines in the region), whereas the corresponding linker region in *Th.* 23B OrfB is not enriched in serine residues. This difference in the AT/ER linker region most likely accounts for a break in the alignment between *Schizochytrium* OrfB and *Th.* 23B OrfB at the start of this region.

Th. 23B OrfB is a 5805 nucleotide sequence (not including the stop codon) which encodes a 1935 amino acid sequence, represented herein as SEQ ID NO:52. The *Th.* 23B OrfB amino acid sequence (SEQ ID NO:52) was compared with known sequences in a standard BLAST search (BLAST parameters: Blastp, low complexity filter Off, program - BLOSUM62,Gap cost - Existence: 11, Extension 1; (BLAST described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety))). At the amino acid level, the sequences with the greatest degree of homology to *Th.* 23B OrfB were *Schizochytrium* Orf B (gb AAK72880.1) (SEQ ID NO:4), over most of OrfB; and *Schizochytrium* OrfC (gb AAK728881.1) (SEQ ID NO:6), over the last domain (the alignment is broken into 2 pieces, as mentioned above). SEQ ID NO:52 first aligns at positions 10 through about 1479 (including the KS, CLF and AT domains) with SEQ ID NO:4 and shows a sequence identity to SEQ ID NO:4 of 52% over 1483 amino acids. SEQ ID NO:52 also aligns at positions 1491 through 1935 (including the ER domain) with SEQ ID NO:6 and shows a sequence identity to SEQ ID NO:4 of 64% over 448 amino acids.

The first domain in the *Th.* 23B OrfB is a KS domain, also referred to herein as *Th.* 23B OrfB-KS. KS domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 1 and about postion 1500 of SEQ ID NO:51 (*Th.* 23B OrfB), represented herein as SEQ ID NO:53. The amino acid sequence containing the *Th.* 23B KS domain is a region of SEQ ID NO: 52

spanning from about position 1 to about position 500 of SEQ ID NO:52, represented herein as SEQ ID NO:54. This region of SEQ ID NO:52 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from about position 1 to about position 450 (positions 1-450 of SEQ ID NO:54). It is noted that the *Th.* 23B OrfB-KS domain contains an active site motif: DXAC*, where C* is the site of acyl group attachment and wherein the C* is at position 201 of SEQ ID NO:52. Also, a characteristic motif at the end of the KS region, GFGG is present in amino acid positions 434-437 of SEQ ID NO:52. The amino acid sequence spanning positions 1-500 of SEQ ID NO:52 is about 64% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 500 amino acids. The amino acid sequence spanning positions 1-450 of SEQ ID NO:52 is about 67% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 442 amino acids.

The second domain in *Th.* 23B OrfB is a CLF domain, also referred to herein as *Th.* 23B OrfB-CLF. CLF domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 1501 and about position 3000 of SEQ ID NO:51 (OrfB), represented herein as SEQ ID NO:55. The amino acid sequence containing the CLF domain is a region of SEQ ID NO: 52 spanning from about position 501 to about position 1000 of SEQ ID NO:52, represented herein as SEQ ID NO:56. This region of SEQ ID NO:52 has a Pfam match to FabB (β-ketoacyl-ACP synthase) spanning from about position 550 to about position 910 (positions 50-410 of SEQ ID NO:56). Although CLF has homology to KS proteins, it lacks an active site cysteine to which the acyl group is attached in KS proteins. The amino acid sequence spanning positions 501-1000 of SEQ ID NO:52 is about 49% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 517 amino acids. The amino acid sequence spanning positions 550-910 of SEQ ID NO:52 is about 54% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 360 amino acids.

The third domain in *Th.* 23B OrfB is an AT domain, also referred to herein as *Th.* 23B OrfB-AT. AT domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 3001 and about position 4500 of SEQ ID NO:51 (*Th.* 23B OrfB), represented herein as SEQ ID NO:58. The amino acid sequence containing the *Th.* 23B AT domain is a region of SEQ ID NO: 52

spanning from about position 1001 to about position 1500 of SEQ ID NO:52, represented herein as SEQ ID NO:58. This region of SEQ ID NO:52 has a Pfam match to FabD (malonyl-CoA:ACP acyltransferase) spanning from about position 1100 to about position 1375 (positions 100-375 of SEQ ID NO:58). Although this AT domain of the PUFA synthases has homology to MAT proteins, it lacks the extended motif of the MAT (key arginine and glutamine residues) and it is not thought to be involved in malonyl-CoA transfers. The GXS*XG motif of acyltransferases is present, with the S* being the site of acyl attachment and located at position 1123 with respect to SEQ ID NO:52. The amino acid sequence spanning positions 1001-1500 of SEQ ID NO:52 is about 44% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 459 amino acids. The amino acid sequence spanning positions 1100-1375 of SEQ ID NO:52 is about 45% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 283 amino acids.

The fourth domain in *Th.* 23B OrfB is an ER domain, also referred to herein as *Th.* 23B OrfB-ER. ER domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 4501 and about position 5805 of SEQ ID NO:51 (OrfB), represented herein as SEQ ID NO:59. The amino acid sequence containing the *Th.* 23B ER domain is a region of SEQ ID NO: 52 spanning from about position 1501 to about position 1935 of SEQ ID NO:52, represented herein as SEQ ID NO:60. This region of SEQ ID NO:52 has a Pfam match to a family of dioxygenases related to 2-nitropropane dioxygenases spanning from about position 1501 to about position 1810 (positions 1-310 of SEQ ID NO:60). That this domain functions as an ER can be further predicted due to homology to a newly characterized ER enzyme from *Streptococcus pneumoniae*. The amino acid sequence spanning positions 1501-1935 of SEQ ID NO:52 is about 66% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 433 amino acids. The amino acid sequence spanning positions 1501-1810 of SEQ ID NO:52 is about 70% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 305 amino acids.

Th. 23B *Open Reading Frame C (OrfC)*:

The complete nucleotide sequence for *Th.* 23B OrfC is represented herein as SEQ ID NO:61. SEQ ID NO:61 encodes the following domains in *Th.* 23B OrfC: (a) two FabA-like

β-hydroxyacyl-ACP dehydrase (DH) domains, both with homology to the FabA protein (an enzyme that catalyzes the synthesis of *trans*-2-decenoyl-ACP and the reversible isomerization of this product to *cis*-3-decenoyl-ACP); and (b) one enoyl-ACP reductase (ER) domain with high homology to the ER domain of *Schizochytrium* OrfB. This domain organization is the same as in *Schizochytrium* Orf C (SEQ ID NO:5).

Th. 23B OrfC is a 4410 nucleotide sequence (not including the stop codon) which encodes a 1470 amino acid sequence, represented herein as SEQ ID NO:62. The *Th.* 23B OrfC amino acid sequence (SEQ ID NO:62) was compared with known sequences in a standard BLAST search (BLAST parameters: Blastp, low complexity filter Off, program - BLOSUM62,Gap cost - Existence: 11, Extension 1; (BLAST described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety))). At the amino acid level, the sequences with the greatest degree of homology to *Th.* 23B OrfC was *Schizochytrium* OrfC (gb AAK728881.1) (SEQ ID NO:6). SEQ ID NO:52 is 66% identical to *Schizochytrium* OrfC (SEQ ID NO:6).

The first domain in *Th.* 23B OrfC is a DH domain, also referred to herein as *Th.* 23B OrfC-DH1. DH domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 1 to about position 1500 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:63. The amino acid sequence containing the *Th.* 23B DH1 domain is a region of SEQ ID NO: 62 spanning from about position 1 to about position 500 of SEQ ID NO:62, represented herein as SEQ ID NO:64. This region of SEQ ID NO:62 has a Pfam match to FabA, as mentioned above, spanning from about position 275 to about position 400 (positions 275-400 of SEQ ID NO:64). The amino acid sequence spanning positions 1-500 of SEQ ID NO:62 is about 66% identical to *Schizochytrium* OrfC (SEQ ID NO:6) over 526 amino acids. The amino acid sequence spanning positions 275-400 of SEQ ID NO:62 is about 81% identical to *Schizochytrium* OrfC (SEQ ID NO:6) over 126 amino acids.

The second domain in *Th.* 23B OrfC is also a DH domain, also referred to herein as *Th.* 23B OrfC-DH2. This is the second of two DH domains in OrfC, and therefore is designated DH2. This domain is contained within the nucleotide sequence spanning from between about position 1501 to about 3000 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:65. The amino acid sequence containing the *Th.* 23B DH2 domain is a region of SEQ ID NO: 62 spanning from about position 501 to about position 1000 of SEQ ID NO:62, represented herein as SEQ ID NO:66. This region of SEQ ID NO:62 has a Pfam match to FabA, as mentioned above, spanning from about position 800 to about position 925 (positions 300-425 of SEQ ID NO:66). The amino acid sequence spanning positions 501-1000 of SEQ ID NO:62 is about 56% identical to *Schizochytrium* OrfC (SEQ ID NO:6) over 518 amino acids. The amino acid sequence spanning positions 800-925 of SEQ ID NO:62 is about 58% identical to *Schizochytrium* OrfC (SEQ ID NO:6) over 124 amino acids.

The third domain in *Th.* 23B OrfC is an ER domain, also referred to herein as *Th.* 23B OrfC-ER. ER domain function has been described in detail above. This domain is contained within the nucleotide sequence spanning from between about position 3001 to about position 4410 of SEQ ID NO:61 (OrfC), represented herein as SEQ ID NO:67. The amino acid sequence containing the *Th.* 23B ER domain is a region of SEQ ID NO: 62 spanning from about position 1001 to about position 1470 of SEQ ID NO:62, represented herein as SEQ ID NO:68. This region of SEQ ID NO:62 has a Pfam match to the dioxygenases related to 2-nitropropane dioxygenases, as mentioned above, spanning from about position 1025 to about position 1320 (positions 25-320 of SEQ ID NO:68). This domain function as an ER can also be predicted due to homology to a newly characterized ER enzyme from *Streptococcus pneumoniae*. The amino acid sequence spanning positions 1001-1470 of SEQ ID NO:62 is about 75% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 474 amino acids. The amino acid sequence spanning positions 1025-1320 of SEQ ID NO:62 is about 81% identical to *Schizochytrium* OrfB (SEQ ID NO:4) over 296 amino acids.

One embodiment of the present invention relates to an isolated protein or domain from a non-bacterial PUFA PKS system, a homologue thereof, and/or a fragment thereof. Also included in the invention are isolated nucleic acid molecules encoding any of the

proteins, domains or peptides described herein (discussed in detail below). According to the present invention, an isolated protein or peptide, such as a protein or peptide from a PUFA PKS system, is a protein or a fragment thereof (including a polypeptide or peptide) that has been removed from its natural milieu (i.e., that has been subject to human manipulation) and can include purified proteins, partially purified proteins, recombinantly produced proteins, and synthetically produced proteins, for example. As such, "isolated" does not reflect the extent to which the protein has been purified. Preferably, an isolated protein of the present invention is produced recombinantly. An isolated peptide can be produced synthetically (e.g., chemically, such as by peptide synthesis) or recombinantly. In addition, and by way of example, a "*Thraustochytrium* PUFA PKS protein" refers to a PUFA PKS protein (generally including a homologue of a naturally occurring PUFA PKS protein) from a *Thraustochytrium* microorganism, or to a PUFA PKS protein that has been otherwise produced from the knowledge of the structure (e.g., sequence), and perhaps the function, of a naturally occurring PUFA PKS protein from *Thraustochytrium*. In other words, general reference to a *Thraustochytrium* PUFA PKS protein includes any PUFA PKS protein that has substantially similar structure and function of a naturally occurring PUFA PKS protein from *Thraustochytrium* or that is a biologically active (i.e., has biological activity) homologue of a naturally occurring PUFA PKS protein from *Thraustochytrium* as described in detail herein. As such, a *Thraustochytrium* PUFA PKS protein can include purified, partially purified, recombinant, mutated/modified and synthetic proteins. The same description applies to reference to other proteins or peptides described herein, such as the PUFA PKS proteins and domains from *Schizochytrium* or from other microorganisms.

According to the present invention, the terms "modification" and "mutation" can be used interchangeably, particularly with regard to the modifications/mutations to the primary amino acid sequences of a protein or peptide (or nucleic acid sequences) described herein. The term "modification" can also be used to describe post-translational modifications to a protein or peptide including, but not limited to, methylation, farnesylation, carboxymethylation, geranyl geranylation, glycosylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, and/or amidation. Modifications can also include,

for example, complexing a protein or peptide with another compound. Such modifications can be considered to be mutations, for example, if the modification is different than the post-translational modification that occurs in the natural, wild-type protein or peptide.

As used herein, the term "homologue" is used to refer to a protein or peptide which differs from a naturally occurring protein or peptide (i.e., the "prototype" or "wild-type" protein) by one or more minor modifications or mutations to the naturally occurring protein or peptide, but which maintains the overall basic protein and side chain structure of the naturally occurring form (i.e., such that the homologue is identifiable as being related to the wild-type protein). Such changes include, but are not limited to: changes in one or a few amino acid side chains; changes one or a few amino acids, including deletions (e.g., a truncated version of the protein or peptide) insertions and/or substitutions; changes in stereochemistry of one or a few atoms; and/or minor derivatizations, including but not limited to: methylation, farnesylation, geranyl geranylation, glycosylation, carboxymethylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, and/or amidation. A homologue can have either enhanced, decreased, or substantially similar properties as compared to the naturally occurring protein or peptide. Preferred homologues of a PUFA PKS protein or domain are described in detail below. It is noted that homologues can include synthetically produced homologues, naturally occurring allelic variants of a given protein or domain, or homologous sequences from organisms other than the organism from which the reference sequence was derived.

Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* (1982) 157: 105-132), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* (1978) 47: 45-148, 1978).

Homologues can be the result of natural allelic variation or natural mutation. A naturally occurring allelic variant of a nucleic acid encoding a protein is a gene that occurs

at essentially the same locus (or loci) in the genome as the gene which encodes such protein, but which, due to natural variations caused by, for example, mutation or recombination, has a similar but not identical sequence. Allelic variants typically encode proteins having similar activity to that of the protein encoded by the gene to which they are being compared. One class of allelic variants can encode the same protein but have different nucleic acid sequences due to the degeneracy of the genetic code. Allelic variants can also comprise alterations in the 5' or 3' untranslated regions of the gene (e.g., in regulatory control regions). Allelic variants are well known to those skilled in the art.

Homologues can be produced using techniques known in the art for the production of proteins including, but not limited to, direct modifications to the isolated, naturally occurring protein, direct protein synthesis, or modifications to the nucleic acid sequence encoding the protein using, for example, classic or recombinant DNA techniques to effect random or targeted mutagenesis.

Modifications or mutations in protein homologues, as compared to the wild-type protein, either increase, decrease, or do not substantially change, the basic biological activity of the homologue as compared to the naturally occurring (wild-type) protein. In general, the biological activity or biological action of a protein refers to any function(s) exhibited or performed by the protein that is ascribed to the naturally occurring form of the protein as measured or observed *in vivo* (i.e., in the natural physiological environment of the protein) or *in vitro* (i.e., under laboratory conditions). Biological activities of PUFA PKS systems and the individual proteins/domains that make up a PUFA PKS system have been described in detail elsewhere herein. Modifications of a protein, such as in a homologue or mimetic (discussed below), may result in proteins having the same biological activity as the naturally occurring protein, or in proteins having decreased or increased biological activity as compared to the naturally occurring protein. Modifications which result in a decrease in protein expression or a decrease in the activity of the protein, can be referred to as inactivation (complete or partial), down-regulation, or decreased action (or activity) of a protein. Similarly, modifications which result in an increase in protein expression or an increase in the activity of the protein, can be referred to as amplification, overproduction,

activation, enhancement, up-regulation or increased action (or activity) of a protein. It is noted that general reference to a homologue having the biological activity of the wild-type protein does not necessarily mean that the homologue has identical biological activity as the wild-type protein, particularly with regard to the level of biological activity. Rather, a homologue can perform the same biological activity as the wild-type protein, but at a reduced or increased level of activity as compared to the wild-type protein. A functional domain of a PUFA PKS system is a domain (i.e., a domain can be a portion of a protein) that is capable of performing a biological function (i.e., has biological activity).

Methods of detecting and measuring PUFA PKS protein or domain biological activity include, but are not limited to, measurement of transcription of a PUFA PKS protein or domain, measurement of translation of a PUFA PKS protein or domain, measurement of posttranslational modification of a PUFA PKS protein or domain, measurement of enzymatic activity of a PUFA PKS protein or domain, and/or measurement production of one or more products of a PUFA PKS system (e.g., PUFA production). It is noted that an isolated protein of the present invention (including a homologue) is not necessarily required to have the biological activity of the wild-type protein. For example, a PUFA PKS protein or domain can be a truncated, mutated or inactive protein, for example. Such proteins are useful in screening assays, for example, or for other purposes such as antibody production. In a preferred embodiment, the isolated proteins of the present invention have biological activity that is similar to that of the wild-type protein (although not necessarily equivalent, as discussed above).

Methods to measure protein expression levels generally include, but are not limited to: Western blot, immunoblot, enzyme-linked immunosorbant assay (ELISA), radioimmunoassay (RIA), immunoprecipitation, surface plasmon resonance, chemiluminescence, fluorescent polarization, phosphorescence, immunohistochemical analysis, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, microcytometry, microarray, microscopy, fluorescence activated cell sorting (FACS), and flow cytometry, as well as assays based on a property of the protein including but not limited to enzymatic activity or interaction with other protein partners. Binding

assays are also well known in the art. For example, a BIAcore machine can be used to determine the binding constant of a complex between two proteins. The dissociation constant for the complex can be determined by monitoring changes in the refractive index with respect to time as buffer is passed over the chip (O'Shannessy et al. Anal. Biochem. 212:457-468 (1993); Schuster et al., Nature 365:343-347 (1993)). Other suitable assays for measuring the binding of one protein to another include, for example, immunoassays such as enzyme linked immunoabsorbent assays (ELISA) and radioimmunoassays (RIA); or determination of binding by monitoring the change in the spectroscopic or optical properties of the proteins through fluorescence, UV absorption, circular dichrosim, or nuclear magnetic resonance (NMR).

In one embodiment, the present invention relates to an isolated protein comprising an amino acid sequence selected from the group consisting of: (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:39, SEQ ID NO:52, SEQ ID NO:62, and biologically active fragments thereof; (b) an amino acid sequence selected from the group consisting of: SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68 and biologically active fragments thereof; (c) an amino acid sequence that is at least about 60% identical to at least 500 consecutive amino acids of the amino acid sequence of (a), wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system; and/or (d) an amino acid sequence that is at least about 60% identical to the amino acid sequence of (b), wherein the amino acid sequence has a biological activity of at least one domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. In a further embodiment, an amino acid sequence including the active site domains or other functional motifs described above for several of the PUFA PKS domains are encompassed by the invention. In one embodiment, the amino acid sequence described above does not include any of the following amino acid sequences: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ

ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32.

In one aspect of the invention, a PUFA PKS protein or domain encompassed by the present invention, including a homologue of a particular PUFA PKS protein or domain described herein, comprises an amino acid sequence that is at least about 60% identical to at least 500 consecutive amino acids of an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:52, or SEQ ID NO:62, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 600 consecutive amino acids, and more preferably to at least about 700 consecutive amino acids, and more preferably to at least about 800 consecutive amino acids, and more preferably to at least about 900 consecutive amino acids, and more preferably to at least about 1000 consecutive amino acids, and more preferably to at least about 1100 consecutive amino acids, and more preferably to at least about 1200 consecutive amino acids, and more preferably to at least about 1300 consecutive amino acids, and more preferably to at least about 1400 consecutive amino acids of any of SEQ ID NO:39, SEQ ID NO:52, or SEQ ID NO:62, or to the full length of SEQ ID NO:62. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 1500 consecutive amino acids, and more preferably to at least about 1600 consecutive amino acids, and more preferably to at least about 1700 consecutive amino acids, and more preferably to at least about 1800 consecutive amino acids, and more preferably to at least about 1900 consecutive amino acids, of any of SEQ ID NO:39 or SEQ ID NO:52, or to the full length of SEQ ID NO:52. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 2000 consecutive amino acids, and more preferably to at least about 2100 consecutive amino acids, and more preferably to at least about 2200 consecutive amino acids, and more preferably to at least about 2300 consecutive amino acids, and more preferably to at least about 2400 consecutive amino acids, and more preferably to at least about 2500 consecutive amino acids, and more preferably to at least about 2600 consecutive amino acids, and more preferably to at least about 2700 consecutive amino acids, and more preferably to at least

about 2800 consecutive amino acids, and even more preferably, to the full length of SEQ ID NO:39. In one embodiment, the amino acid sequence described above does not include any of the following amino acid sequences: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32.

In another aspect, a PUFA PKS protein or domain encompassed by the present invention, including homologues as described above, comprises an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:52, or SEQ ID NO:62, over any of the consecutive amino acid lengths described in the paragraph above, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In one embodiment, the amino acid sequence described above does not include any of the following amino acid sequences: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32.

In one aspect of the invention, a PUFA PKS protein or domain encompassed by the present invention, including a homologue as described above, comprises an amino acid sequence that is at least about 60% identical to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the protein is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and

more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In one embodiment, the amino acid sequence described above does not include any of the following amino acid sequences: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32.

In another aspect, a PUFA PKS protein or domain encompassed by the present invention, including a homologue as described above, comprises an amino acid sequence that is at least about 50% identical to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In another aspect, the amino acid sequence of the protein is at least about 55% identical, and more preferably at least about 60% identical, to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the protein is at least about 65% identical to an amino acid sequence chosen from SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56 and SEQ ID NO:58, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In another aspect, the amino acid sequence of the protein is at least about 70% identical, and more preferably at least about 75% identical, to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50,

SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, and SEQ ID NO:64, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In another aspect, the amino acid sequence of the protein is at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical, to an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In one embodiment, the amino acid sequence described above does not include any of the following amino acid sequences: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32.

In a preferred embodiment an isolated protein or domain of the present invention comprises, consists essentially of, or consists of, an amino acid sequence chosen from: SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, or any biologically active fragments thereof, including any fragments that have a biological activity of at least one domain of a PUFA PKS system.

In one aspect of the present invention, the following *Schizochytrium* proteins and domains are useful in one or more embodiments of the present invention, all of which have been previously described in detail in U.S. Patent Application Serial No. 10/124,800, *supra*. In one aspect of the invention, a PUFA PKS protein or domain useful in the present invention comprises an amino acid sequence that is at least about 60% identical to at least 500 consecutive amino acids of an amino acid sequence chosen from: SEQ ID NO:2, SEQ

ID NO:4, and SEQ ID NO:6; wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 600 consecutive amino acids, and more preferably to at least about 700 consecutive amino acids, and more preferably to at least about 800 consecutive amino acids, and more preferably to at least about 900 consecutive amino acids, and more preferably to at least about 1000 consecutive amino acids, and more preferably to at least about 1100 consecutive amino acids, and more preferably to at least about 1200 consecutive amino acids, and more preferably to at least about 1300 consecutive amino acids, and more preferably to at least about 1400 consecutive amino acids, and more preferably to at least about 1500 consecutive amino acids of any of SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:6, or to the full length of SEQ ID NO:6. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 1600 consecutive amino acids, and more preferably to at least about 1700 consecutive amino acids, and more preferably to at least about 1800 consecutive amino acids, and more preferably to at least about 1900 consecutive amino acids, and more preferably to at least about 2000 consecutive amino acids of any of SEQ ID NO:2 or SEQ ID NO:4, or to the full length of SEQ ID NO:4. In a further aspect, the amino acid sequence of the protein is at least about 60% identical to at least about 2100 consecutive amino acids, and more preferably to at least about 2200 consecutive amino acids, and more preferably to at least about 2300 consecutive amino acids, and more preferably to at least about 2400 consecutive amino acids, and more preferably to at least about 2500 consecutive amino acids, and more preferably to at least about 2600 consecutive amino acids, and more preferably to at least about 2700 consecutive amino acids, and more preferably to at least about 2800 consecutive amino acids, and even more preferably, to the full length of SEQ ID NO:2.

In another aspect, a PUFA PKS protein or domain useful in one or more embodiments of the present invention comprises an amino acid sequence that is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more

preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to an amino acid sequence chosen from: SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, over any of the consecutive amino acid lengths described in the paragraph above, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system.

In another aspect of the invention, a PUFA PKS protein or domain useful in one or more embodiments of the present invention comprises an amino acid sequence that is at least about 60% identical to an amino acid sequence chosen from: SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:32, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system. In a further aspect, the amino acid sequence of the protein is at least about 65% identical, and more preferably at least about 70% identical, and more preferably at least about 75% identical, and more preferably at least about 80% identical, and more preferably at least about 85% identical, and more preferably at least about 90% identical, and more preferably at least about 95% identical, and more preferably at least about 96% identical, and more preferably at least about 97% identical, and more preferably at least about 98% identical, and more preferably at least about 99% identical to an amino acid sequence chosen from: SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, wherein the amino acid sequence has a biological activity of at least one domain of a PUFA PKS system.

In yet another aspect of the invention, a PUFA PKS protein or domain useful in one or more embodiments of the present invention comprises, consists essentially of, or consists of, an amino acid sequence chosen from: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 or any biologically active fragments thereof, including any fragments that have a biological activity of at least one domain of a PUFA PKS system.

According to the present invention, the term "contiguous" or "consecutive", with regard to nucleic acid or amino acid sequences described herein, means to be connected in an unbroken sequence. For example, for a first sequence to comprise 30 contiguous (or consecutive) amino acids of a second sequence, means that the first sequence includes an unbroken sequence of 30 amino acid residues that is 100% identical to an unbroken sequence of 30 amino acid residues in the second sequence. Similarly, for a first sequence to have "100% identity" with a second sequence means that the first sequence exactly matches the second sequence with no gaps between nucleotides or amino acids.

As used herein, unless otherwise specified, reference to a percent (%) identity refers to an evaluation of homology which is performed using: (1) a BLAST 2.0 Basic BLAST homology search using blastp for amino acid searches, blastn for nucleic acid searches, and blastX for nucleic acid searches and searches of translated amino acids in all 6 open reading frames, all with standard default parameters, wherein the query sequence is filtered for low complexity regions by default (described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety); (2) a BLAST 2 alignment (using the parameters described below); (3) and/or PSI-BLAST with the standard default parameters (Position-Specific Iterated BLAST). It is noted that due to some differences in the standard parameters between BLAST 2.0 Basic BLAST and BLAST 2, two specific sequences might be recognized as having significant homology using the BLAST 2 program, whereas a search performed in BLAST 2.0 Basic BLAST using one of the sequences as the query sequence may not identify the second sequence in the top matches. In addition, PSI-BLAST provides an automated, easy-to-use version of a "profile" search, which is a sensitive way to look for sequence homologues. The program first performs a gapped BLAST database search. The PSI-BLAST program uses the information from any significant alignments returned to construct a position-specific score matrix, which replaces the query sequence for the next round of database searching. Therefore, it is to be understood that percent identity can be determined by using any one of these programs.

Two specific sequences can be aligned to one another using BLAST 2 sequence as described in Tatusova and Madden, (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* **174**, 247, incorporated herein by reference in its entirety. BLAST 2 sequence alignment is performed in blastp or blastn using the BLAST 2.0 algorithm to perform a Gapped BLAST search (BLAST 2.0) between the two sequences allowing for the introduction of gaps (deletions and insertions) in the resulting alignment. For purposes of clarity herein, a BLAST 2 sequence alignment is performed using the standard default parameters as follows.

For blastn, using 0 BLOSUM62 matrix:

Reward for match = 1
Penalty for mismatch = -2
Open gap (5) and extension gap (2) penalties
gap x_dropoff (50) expect (10) word size (11) filter (on)

For blastp, using 0 BLOSUM62 matrix:
Open gap (11) and extension gap (1) penalties
gap x_dropoff (50) expect (10) word size (3) filter (on).

According to the present invention, an amino acid sequence that has a biological activity of at least one domain of a PUFA PKS system is an amino acid sequence that has the biological activity of at least one domain of the PUFA PKS system described in detail herein, as previously exemplified by the *Schizochytrium* PUFA PKS system or as additionally exemplified herein by the *Thraustochytrium* PUFA PKS system. The biological activities of the various domains within the *Schizochytrium* or *Thraustochytrium* PUFA PKS systems have been described in detail above. Therefore, an isolated protein useful in the present invention can include the translation product of any PUFA PKS open reading frame, any PUFA PKS domain, biologically active fragment thereof, or any homologue of a naturally occurring PUFA PKS open reading frame product or domain which has biological activity.

In another embodiment of the invention, an amino acid sequence having the biological activity of at least one domain of a PUFA PKS system of the present invention includes an amino acid sequence that is sufficiently similar to a naturally occurring PUFA PKS protein or polypeptide that a nucleic acid sequence encoding the amino acid sequence

is capable of hybridizing under moderate, high, or very high stringency conditions (described below) to (i.e., with) a nucleic acid molecule encoding the naturally occurring PUFA PKS protein or polypeptide (i.e., to the complement of the nucleic acid strand encoding the naturally occurring PUFA PKS protein or polypeptide). Preferably, an amino acid sequence having the biological activity of at least one domain of a PUFA PKS system of the present invention is encoded by a nucleic acid sequence that hybridizes under moderate, high or very high stringency conditions to the complement of a nucleic acid sequence that encodes any of the above-described amino acid sequences for a PUFA PKS protein or domain. Methods to deduce a complementary sequence are known to those skilled in the art. It should be noted that since amino acid sequencing and nucleic acid sequencing technologies are not entirely error-free, the sequences presented herein, at best, represent apparent sequences of PUFA PKS domains and proteins of the present invention.

As used herein, hybridization conditions refer to standard hybridization conditions under which nucleic acid molecules are used to identify similar nucleic acid molecules. Such standard conditions are disclosed, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety (see specifically, pages 9.31-9.62). In addition, formulae to calculate the appropriate hybridization and wash conditions to achieve hybridization permitting varying degrees of mismatch of nucleotides are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284; Meinkoth et al., *ibid.*, is incorporated by reference herein in its entirety.

More particularly, moderate stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 70% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 30% or less mismatch of nucleotides). High stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 80% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 20% or less mismatch of

nucleotides). Very high stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 90% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 10% or less mismatch of nucleotides). As discussed above, one of skill in the art can use the formulae in Meinkoth et al., *ibid.* to calculate the appropriate hybridization and wash conditions to achieve these particular levels of nucleotide mismatch. Such conditions will vary, depending on whether DNA:RNA or DNA:DNA hybrids are being formed. Calculated melting temperatures for DNA:DNA hybrids are 10°C less than for DNA:RNA hybrids. In particular embodiments, stringent hybridization conditions for DNA:DNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 20°C and about 35°C (lower stringency), more preferably, between about 28°C and about 40°C (more stringent), and even more preferably, between about 35°C and about 45°C (even more stringent), with appropriate wash conditions. In particular embodiments, stringent hybridization conditions for DNA:RNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 30°C and about 45°C, more preferably, between about 38°C and about 50°C, and even more preferably, between about 45°C and about 55°C, with similarly stringent wash conditions. These values are based on calculations of a melting temperature for molecules larger than about 100 nucleotides, 0% formamide and a G + C content of about 40%. Alternatively, T_m can be calculated empirically as set forth in Sambrook et al., *supra*, pages 9.31 to 9.62. In general, the wash conditions should be as stringent as possible, and should be appropriate for the chosen hybridization conditions. For example, hybridization conditions can include a combination of salt and temperature conditions that are approximately 20-25°C below the calculated T_m of a particular hybrid, and wash conditions typically include a combination of salt and temperature conditions that are approximately 12-20°C below the calculated T_m of the particular hybrid. One example of hybridization conditions suitable for use with DNA:DNA hybrids includes a 2-24 hour hybridization in 6X SSC (50% formamide) at about 42°C, followed by washing steps that include one or more washes at room temperature in about 2X SSC, followed by additional

washes at higher temperatures and lower ionic strength (e.g., at least one wash as about 37°C in about 0.1X-0.5X SSC, followed by at least one wash at about 68°C in about 0.1X-0.5X SSC).

The present invention also includes a fusion protein that includes any PUFA PKS protein or domain or any homologue or fragment thereof attached to one or more fusion segments. Suitable fusion segments for use with the present invention include, but are not limited to, segments that can: enhance a protein's stability; provide other desirable biological activity; and/or assist with the purification of the protein (e.g., by affinity chromatography). A suitable fusion segment can be a domain of any size that has the desired function (e.g., imparts increased stability, solubility, biological activity; and/or simplifies purification of a protein). Fusion segments can be joined to amino and/or carboxyl termini of the protein and can be susceptible to cleavage in order to enable straight-forward recovery of the desired protein. Fusion proteins are preferably produced by culturing a recombinant cell transfected with a fusion nucleic acid molecule that encodes a protein including the fusion segment attached to either the carboxyl and/or amino terminal end of the protein of the invention as discussed above.

In one embodiment of the present invention, any of the above-described PUFA PKS amino acid sequences, as well as homologues of such sequences, can be produced with from at least one, and up to about 20, additional heterologous amino acids flanking each of the C- and/or N-terminal end of the given amino acid sequence. The resulting protein or polypeptide can be referred to as "consisting essentially of" a given amino acid sequence. According to the present invention, the heterologous amino acids are a sequence of amino acids that are not naturally found (i.e., not found in nature, *in vivo*) flanking the given amino acid sequence or which would not be encoded by the nucleotides that flank the naturally occurring nucleic acid sequence encoding the given amino acid sequence as it occurs in the gene, if such nucleotides in the naturally occurring sequence were translated using standard codon usage for the organism from which the given amino acid sequence is derived. Similarly, the phrase "consisting essentially of", when used with reference to a nucleic acid sequence herein, refers to a nucleic acid sequence encoding a given amino acid sequence that

can be flanked by from at least one, and up to as many as about 60, additional heterologous nucleotides at each of the 5' and/or the 3' end of the nucleic acid sequence encoding the given amino acid sequence. The heterologous nucleotides are not naturally found (i.e., not found in nature, *in vivo*) flanking the nucleic acid sequence encoding the given amino acid sequence as it occurs in the natural gene.

The minimum size of a protein or domain and/or a homologue or fragment thereof of the present invention is, in one aspect, a size sufficient to have the requisite biological activity, or sufficient to serve as an antigen for the generation of an antibody or as a target in an *in vitro* assay. In one embodiment, a protein of the present invention is at least about 8 amino acids in length (e.g., suitable for an antibody epitope or as a detectable peptide in an assay), or at least about 25 amino acids in length, or at least about 50 amino acids in length, or at least about 100 amino acids in length, or at least about 150 amino acids in length, or at least about 200 amino acids in length, or at least about 250 amino acids in length, or at least about 300 amino acids in length, or at least about 350 amino acids in length, or at least about 400 amino acids in length, or at least about 450 amino acids in length, or at least about 500 amino acids in length, or at least about 750 amino acids in length, and so on, in any length between 8 amino acids and up to the full length of a protein or domain of the invention or longer, in whole integers (e.g., 8, 9, 10,...25, 26,...500, 501,...1234, 1235,...). There is no limit, other than a practical limit, on the maximum size of such a protein in that the protein can include a portion of a PUFA PKS protein, domain, or biologically active or useful fragment thereof, or a full-length PUFA PKS protein or domain, plus additional sequence (e.g., a fusion protein sequence), if desired.

Further embodiments of the present invention include isolated nucleic acid molecules comprising, consisting essentially of, or consisting of nucleic acid sequences that encode any of the above-identified proteins or domains, including a homologue or fragment thereof, as well as nucleic acid sequences that are fully complementary thereto. In accordance with the present invention, an isolated nucleic acid molecule is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found

in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. An isolated nucleic acid molecule can include a gene. An isolated nucleic acid molecule that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes naturally found on the same chromosome. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecule can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Although the phrase "nucleic acid molecule" primarily refers to the physical nucleic acid molecule and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the nucleic acid molecule, the two phrases can be used interchangeably, especially with respect to a nucleic acid molecule, or a nucleic acid sequence, being capable of encoding a protein or domain of a protein.

Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. Isolated nucleic acid molecules include natural nucleic acid molecules and homologues thereof, including, but not limited to, natural allelic variants and modified nucleic acid molecules in which nucleotides have been inserted, deleted, substituted, and/or inverted in such a manner that such modifications provide the desired effect on PUFA PKS system biological activity as described herein. Protein homologues (e.g., proteins encoded by nucleic acid homologues) have been discussed in detail above.

A nucleic acid molecule homologue can be produced using a number of methods known to those skilled in the art (see, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989). For example, nucleic acid molecules can be modified using a variety of techniques including, but not limited to, classic mutagenesis techniques and recombinant DNA techniques, such as site-directed mutagenesis,

chemical treatment of a nucleic acid molecule to induce mutations, restriction enzyme cleavage of a nucleic acid fragment, ligation of nucleic acid fragments, PCR amplification and/or mutagenesis of selected regions of a nucleic acid sequence, synthesis of oligonucleotide mixtures and ligation of mixture groups to "build" a mixture of nucleic acid molecules and combinations thereof. Nucleic acid molecule homologues can be selected from a mixture of modified nucleic acids by screening for the function of the protein encoded by the nucleic acid and/or by hybridization with a wild-type gene.

The minimum size of a nucleic acid molecule of the present invention is a size sufficient to form a probe or oligonucleotide primer that is capable of forming a stable hybrid (e.g., under moderate, high or very high stringency conditions) with the complementary sequence of a nucleic acid molecule useful in the present invention, or of a size sufficient to encode an amino acid sequence having a biological activity of at least one domain of a PUFA PKS system according to the present invention. As such, the size of the nucleic acid molecule encoding such a protein can be dependent on nucleic acid composition and percent homology or identity between the nucleic acid molecule and complementary sequence as well as upon hybridization conditions *per se* (e.g., temperature, salt concentration, and formamide concentration). The minimal size of a nucleic acid molecule that is used as an oligonucleotide primer or as a probe is typically at least about 12 to about 15 nucleotides in length if the nucleic acid molecules are GC-rich and at least about 15 to about 18 bases in length if they are AT-rich. There is no limit, other than a practical limit, on the maximal size of a nucleic acid molecule of the present invention, in that the nucleic acid molecule can include a sequence sufficient to encode a biologically active fragment of a domain of a PUFA PKS system, an entire domain of a PUFA PKS system, several domains within an open reading frame (Orf) of a PUFA PKS system, an entire Orf of a PUFA PKS system, or more than one Orf of a PUFA PKS system.

In one embodiment of the present invention, an isolated nucleic acid molecule comprises, consists essentially of, or consists of a nucleic acid sequence encoding any of the above-described amino acid sequences, including any of the amino acid sequences, or homologues thereof, from a *Schizochytrium* or *Thraustochytrium* described herein. In one

aspect, the nucleic acid sequence is selected from the group of: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, or SEQ ID NO:67, or homologues (including sequences that are at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to such sequences), or fragments thereof, or any complementary sequences thereof.

Another embodiment of the present invention includes a recombinant nucleic acid molecule comprising a recombinant vector and a nucleic acid sequence encoding protein or peptide having a biological activity of at least one domain (or homologue or fragment thereof) of a PUFA PKS system as described herein. Such nucleic acid sequences are described in detail above. According to the present invention, a recombinant vector is an engineered (i.e., artificially produced) nucleic acid molecule that is used as a tool for manipulating a nucleic acid sequence of choice and for introducing such a nucleic acid sequence into a host cell. The recombinant vector is therefore suitable for use in cloning, sequencing, and/or otherwise manipulating the nucleic acid sequence of choice, such as by expressing and/or delivering the nucleic acid sequence of choice into a host cell to form a recombinant cell. Such a vector typically contains heterologous nucleic acid sequences, that is nucleic acid sequences that are not naturally found adjacent to nucleic acid sequence to be cloned or delivered, although the vector can also contain regulatory nucleic acid sequences (e.g., promoters, untranslated regions) which are naturally found adjacent to nucleic acid molecules of the present invention or which are useful for expression of the nucleic acid molecules of the present invention (discussed in detail below). The vector can be either RNA or DNA, either prokaryotic or eukaryotic, and typically is a plasmid. The vector can be maintained as an extrachromosomal element (e.g., a plasmid) or it can be integrated into the chromosome of a recombinant organism (e.g., a microbe or a plant). The entire vector can remain in place within a host cell, or under certain conditions, the plasmid DNA can be

deleted, leaving behind the nucleic acid molecule of the present invention. The integrated nucleic acid molecule can be under chromosomal promoter control, under native or plasmid promoter control, or under a combination of several promoter controls. Single or multiple copies of the nucleic acid molecule can be integrated into the chromosome. A recombinant vector of the present invention can contain at least one selectable marker.

In one embodiment, a recombinant vector used in a recombinant nucleic acid molecule of the present invention is an expression vector. As used herein, the phrase "expression vector" is used to refer to a vector that is suitable for production of an encoded product (e.g., a protein of interest). In this embodiment, a nucleic acid sequence encoding the product to be produced (e.g., a PUFA PKS domain) is inserted into the recombinant vector to produce a recombinant nucleic acid molecule. The nucleic acid sequence encoding the protein to be produced is inserted into the vector in a manner that operatively links the nucleic acid sequence to regulatory sequences in the vector which enable the transcription and translation of the nucleic acid sequence within the recombinant host cell.

In another embodiment, a recombinant vector used in a recombinant nucleic acid molecule of the present invention is a targeting vector. As used herein, the phrase "targeting vector" is used to refer to a vector that is used to deliver a particular nucleic acid molecule into a recombinant host cell, wherein the nucleic acid molecule is used to delete or inactivate an endogenous gene within the host cell or microorganism (i.e., used for targeted gene disruption or knock-out technology). Such a vector may also be known in the art as a "knock-out" vector. In one aspect of this embodiment, a portion of the vector, but more typically, the nucleic acid molecule inserted into the vector (i.e., the insert), has a nucleic acid sequence that is homologous to a nucleic acid sequence of a target gene in the host cell (i.e., a gene which is targeted to be deleted or inactivated). The nucleic acid sequence of the vector insert is designed to bind to the target gene such that the target gene and the insert undergo homologous recombination, whereby the endogenous target gene is deleted, inactivated or attenuated (i.e., by at least a portion of the endogenous target gene being mutated or deleted). The use of this type of recombinant vector to replace an endogenous *Schizochytrium* gene with a recombinant gene is described in the Examples section, and the

general technique for genetic transformation of Thraustochytrids is described in detail in U.S. Patent Application Serial No. 10/124,807, published as U.S. Patent Application Publication No. 20030166207, published September 4, 2003.

Typically, a recombinant nucleic acid molecule includes at least one nucleic acid molecule of the present invention operatively linked to one or more expression control sequences. As used herein, the phrase "recombinant molecule" or "recombinant nucleic acid molecule" primarily refers to a nucleic acid molecule or nucleic acid sequence operatively linked to a expression control sequence, but can be used interchangeably with the phrase "nucleic acid molecule", when such nucleic acid molecule is a recombinant molecule as discussed herein. According to the present invention, the phrase "operatively linked" refers to linking a nucleic acid molecule to an expression control sequence (e.g., a transcription control sequence and/or a translation control sequence) in a manner such that the molecule is able to be expressed when transfected (i.e., transformed, transduced, transfected, conjugated or conduced) into a host cell. Transcription control sequences are sequences which control the initiation, elongation, or termination of transcription. Particularly important transcription control sequences are those which control transcription initiation, such as promoter, enhancer, operator and repressor sequences. Suitable transcription control sequences include any transcription control sequence that can function in a host cell or organism into which the recombinant nucleic acid molecule is to be introduced.

Recombinant nucleic acid molecules of the present invention can also contain additional regulatory sequences, such as translation regulatory sequences, origins of replication, and other regulatory sequences that are compatible with the recombinant cell. In one embodiment, a recombinant molecule of the present invention, including those which are integrated into the host cell chromosome, also contains secretory signals (i.e., signal segment nucleic acid sequences) to enable an expressed protein to be secreted from the cell that produces the protein. Suitable signal segments include a signal segment that is naturally associated with the protein to be expressed or any heterologous signal segment capable of directing the secretion of the protein according to the present invention. In another embodiment, a recombinant molecule of the present invention comprises a leader sequence

to enable an expressed protein to be delivered to and inserted into the membrane of a host cell. Suitable leader sequences include a leader sequence that is naturally associated with the protein, or any heterologous leader sequence capable of directing the delivery and insertion of the protein to the membrane of a cell.

The present inventors have found that the *Schizochytrium* PUFA PKS Orfs A and B are closely linked in the genome and region between the Orfs has been sequenced. The Orfs are oriented in opposite directions and 4244 base pairs separate the start (ATG) codons (i.e. they are arranged as follows: 3'OrfA5' - 4244 bp - 5'OrfB3'). Examination of the 4244 bp intergenic region did not reveal any obvious Orfs (no significant matches were found on a BlastX search). Both Orfs A and B are highly expressed in *Schizochytrium*, at least during the time of oil production, implying that active promoter elements are embedded in this intergenic region. These genetic elements are believed to have utility as a bi-directional promoter sequence for transgenic applications. For example, in a preferred embodiment, one could clone this region, place any genes of interest at each end and introduce the construct into *Schizochytrium* (or some other host in which the promoters can be shown to function). It is predicted that the regulatory elements, under the appropriate conditions, would provide for coordinated, high level expression of the two introduced genes. The complete nucleotide sequence for the regulatory region containing *Schizochytrium* PUFA PKS regulatory elements (e.g., a promoter) is represented herein as SEQ ID NO:36.

In a similar manner, OrfC is highly expressed in *Schizochytrium* during the time of oil production and regulatory elements are expected to reside in the region upstream of its start codon. A region of genomic DNA upstream of OrfC has been cloned and sequenced and is represented herein as (SEQ ID NO:37). This sequence contains the 3886 nt immediately upstream of the OrfC start codon. Examination of this region did not reveal any obvious Orfs (i.e., no significant matches were found on a BlastX search). It is believed that regulatory elements contained in this region, under the appropriate conditions, will provide for high-level expression of a gene placed behind them. Additionally, under the appropriate conditions, the level of expression may be coordinated with genes under control of the A - B intergenic region (SEQ ID NO:36).

Therefore, in one embodiment, a recombinant nucleic acid molecule useful in the present invention, as disclosed herein, can include a PUFA PKS regulatory region contained within SEQ ID NO:36 and/or SEQ ID NO:37. Such a regulatory region can include any portion (fragment) of SEQ ID NO:36 and/or SEQ ID NO:37 that has at least basal PUFA PKS transcriptional activity.

One or more recombinant molecules of the present invention can be used to produce an encoded product (e.g., a PUFA PKS domain, protein, or system) of the present invention. In one embodiment, an encoded product is produced by expressing a nucleic acid molecule as described herein under conditions effective to produce the protein. A preferred method to produce an encoded protein is by transfecting a host cell with one or more recombinant molecules to form a recombinant cell. Suitable host cells to transfect include, but are not limited to, any bacterial, fungal (e.g., yeast), insect, plant or animal cell that can be transfected. In one embodiment of the invention, a preferred host cell is a Thraustochytrid host cell (described in detail below) or a plant host cell. Host cells can be either untransfected cells or cells that are already transfected with at least one other recombinant nucleic acid molecule.

According to the present invention, the term "transfection" is used to refer to any method by which an exogenous nucleic acid molecule (i.e., a recombinant nucleic acid molecule) can be inserted into a cell. The term "transformation" can be used interchangeably with the term "transfection" when such term is used to refer to the introduction of nucleic acid molecules into microbial cells, such as algae, bacteria and yeast, or into plants. In microbial systems, the term "transformation" is used to describe an inherited change due to the acquisition of exogenous nucleic acids by the microorganism or plant and is essentially synonymous with the term "transfection." However, in animal cells, transformation has acquired a second meaning which can refer to changes in the growth properties of cells in culture after they become cancerous, for example. Therefore, to avoid confusion, the term "transfection" is preferably used with regard to the introduction of exogenous nucleic acids into animal cells, and the term "transfection" will be used herein to generally encompass transfection of animal cells, and transformation of microbial cells or plant cells, to the extent

that the terms pertain to the introduction of exogenous nucleic acids into a cell. Therefore, transfection techniques include, but are not limited to, transformation, particle bombardment, diffusion, active transport, bath sonication, electroporation, microinjection, lipofection, adsorption, infection and protoplast fusion.

It will be appreciated by one skilled in the art that use of recombinant DNA technologies can improve control of expression of transfected nucleic acid molecules by manipulating, for example, the number of copies of the nucleic acid molecules within the host cell, the efficiency with which those nucleic acid molecules are transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Additionally, the promoter sequence might be genetically engineered to improve the level of expression as compared to the native promoter. Recombinant techniques useful for controlling the expression of nucleic acid molecules include, but are not limited to, integration of the nucleic acid molecules into one or more host cell chromosomes, addition of vector stability sequences to plasmids, substitutions or modifications of transcription control signals (e.g., promoters, operators, enhancers), substitutions or modifications of translational control signals (e.g., ribosome binding sites, Shine-Dalgarno sequences), modification of nucleic acid molecules to correspond to the codon usage of the host cell, and deletion of sequences that destabilize transcripts.

General discussion above with regard to recombinant nucleic acid molecules and transfection of host cells is intended to be applied to any recombinant nucleic acid molecule discussed herein, including those encoding any amino acid sequence having a biological activity of at least one domain from a PUFA PKS, those encoding amino acid sequences from other PKS systems, and those encoding other proteins or domains.

Polyunsaturated fatty acids (PUFAs) are essential membrane components in higher eukaryotes and the precursors of many lipid-derived signaling molecules. The PUFA PKS system of the present invention uses pathways for PUFA synthesis that do not require desaturation and elongation of saturated fatty acids. The pathways catalyzed by PUFA PKSs that are distinct from previously recognized PKSs in both structure and mechanism.

Generation of *cis* double bonds is suggested to involve position-specific isomerases; these enzymes are believed to be useful in the production of new families of antibiotics.

To produce significantly high yields of one or more desired polyunsaturated fatty acids or other bioactive molecules, an organism, preferably a microorganism or a plant, and most preferably a Thraustochytrid microorganism, can be genetically modified to alter the activity and particularly, the end product, of the PUFA PKS system in the microorganism or plant.

Therefore, one embodiment of the present invention relates to a genetically modified microorganism, wherein the microorganism expresses a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. The domain of the PUFA PKS system can include any of the domains, including homologues thereof, for PUFA PKS systems as described above (e.g., for *Schizochytrium* and *Thraustochytrium*), and can also include any domain of a PUFA PKS system from any other non-bacterial microorganism, including any eukaryotic microorganism, including any Thraustochytrid microorganism or any domain of a PUFA PKS system from a microorganism identified by a screening method as described in U.S. Patent Application Serial No. 10/124,800, *supra*. The genetic modification affects the activity of the PKS system in the organism. The screening process described in U.S. Patent Application Serial No. 10/124,800 includes the steps of: (a) selecting a microorganism that produces at least one PUFA; and, (b) identifying a microorganism from (a) that has an ability to produce increased PUFAs under dissolved oxygen conditions of less than about 5% of saturation in the fermentation medium, as compared to production of PUFAs by the microorganism under dissolved oxygen conditions of greater than about 5% of saturation, and preferably about 10%, and more preferably about 15%, and more preferably about 20% of saturation in the fermentation medium.

In one aspect, such an organism can endogenously contain and express a PUFA PKS system, and the genetic modification can be a genetic modification of one or more of the functional domains of the endogenous PUFA PKS system, whereby the modification has some effect on the activity of the PUFA PKS system. In another aspect, such an organism

can endogenously contain and express a PUFA PKS system, and the genetic modification can be an introduction of at least one exogenous nucleic acid sequence (e.g., a recombinant nucleic acid molecule), wherein the exogenous nucleic acid sequence encodes at least one biologically active domain or protein from a second PKS system and/or a protein that affects the activity of the PUFA PKS system (e.g., a phosphopantetheinyl transferases (PPTase), discussed below). In yet another aspect, the organism does not necessarily endogenously (naturally) contain a PUFA PKS system, but is genetically modified to introduce at least one recombinant nucleic acid molecule encoding an amino acid sequence having the biological activity of at least one domain of a PUFA PKS system. In this aspect, PUFA PKS activity is affected by introducing or increasing PUFA PKS activity in the organism. Various embodiments associated with each of these aspects will be discussed in greater detail below.

It is to be understood that a genetic modification of a PUFA PKS system or an organism comprising a PUFA PKS system can involve the modification of at least one domain of a PUFA PKS system (including a portion of a domain), more than one or several domains of a PUFA PKS system (including adjacent domains, non-contiguous domains, or domains on different proteins in the PUFA PKS system), entire proteins of the PUFA PKS system, and the entire PUFA PKS system (e.g., all of the proteins encoded by the PUFA PKS genes). As such, modifications can include a small modification to a single domain of an endogenous PUFA PKS system; to substitution, deletion or addition to one or more domains or proteins of a given PUFA PKS system; up to replacement of the entire PUFA PKS system in an organism with the PUFA PKS system from a different organism. One of skill in the art will understand that any genetic modification to a PUFA PKS system is encompassed by the invention.

As used herein, a genetically modified microorganism can include a genetically modified bacterium, protist, microalgae, fungus, or other microbe, and particularly, any of the genera of the order Thraustochytriales (e.g., a Thraustochytrid) described herein (e.g., *Schizochytrium*, *Thraustochytrium*, *Japonochytrium*, *Labyrinthula*, *Labyrinthuloides*, etc.). Such a genetically modified microorganism has a genome which is modified (i.e., mutated or changed) from its normal (i.e., wild-type or naturally occurring) form such that the desired

result is achieved (i.e., increased or modified PUFA PKS activity and/or production of a desired product using the PKS system). Genetic modification of a microorganism can be accomplished using classical strain development and/or molecular genetic techniques. Such techniques known in the art and are generally disclosed for microorganisms, for example, in Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press. The reference Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety. A genetically modified microorganism can include a microorganism in which nucleic acid molecules have been inserted, deleted or modified (i.e., mutated; e.g., by insertion, deletion, substitution, and/or inversion of nucleotides), in such a manner that such modifications provide the desired effect within the microorganism.

Preferred microorganism host cells to modify according to the present invention include, but are not limited to, any bacteria, protist, microalga, fungus, or protozoa. In one aspect, preferred microorganisms to genetically modify include, but are not limited to, any microorganism of the order Thraustochytriales, including any microorganism in the families Thraustochytriaceae and Labyrinthulaceae. Particularly preferred host cells for use in the present invention could include microorganisms from a genus including, but not limited to: *Thraustochytrium, , Japonochytrium, Aplanochytrium, Elina* and *Schizochytrium* within the Thraustochytriaceae and *Labyrinthula, Labyrinthuloides*, and *Labyrinthomyxa* within the Labyrinthulaceae. Preferred species within these genera include, but are not limited to: any species within *Labyrinthula*, including *Labrinthula* sp., *Labyrinthula algeriensis, Labyrinthula cienkowskii, Labyrinthula chattonii, Labyrinthula coenocystis, Labyrinthula macrocystis, Labyrinthula macrocystis atlantica, Labyrinthula macrocystis macrocystis, Labyrinthula magnifica, Labyrinthula minuta, Labyrinthula roscoffensis, Labyrinthula valkanovii, Labyrinthula vitellina, Labyrinthula vitellina pacifica, Labyrinthula vitellina vitellina, Labyrinthula zopfii*; any *Labyrinthuloides* species, including *Labyrinthuloides* sp., *Labyrinthuloides minuta, Labyrinthuloides schizochytrops*; any *Labyrinthomyxa* species, including *Labyrinthomyxa* sp., *Labyrinthomyxa pohlia, Labyrinthomyxa sauvageaui*, any *Aplanochytrium* species, including *Aplanochytrium* sp. and *Aplanochytrium kerguelensis*; any *Elina* species, including *Elina* sp., *Elina marisalba, Elina sinorifica*; any

Japanochytrium species, including *Japanochytrium* sp., *Japanochytrium marinum*; any *Schizochytrium* species, including *Schizochytrium* sp., *Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum, Schizochytrium octosporum;* and any *Thraustochytrium* species, including *Thraustochytrium* sp., *Thraustochytrium aggregatum, Thraustochytrium arudimentale, Thraustochytrium aureum, Thraustochytrium benthicola, Thraustochytrium globosum, Thraustochytrium kinnei, Thraustochytrium motivum, Thraustochytrium pachydermum, Thraustochytrium proliferum, Thraustochytrium roseum, Thraustochytrium striatum, Ulkenia* sp., *Ulkenia minuta, Ulkenia profunda, Ulkenia radiate, Ulkenia sarkariana,* and *Ulkenia visurgensis.* Particularly preferred species within these genera include, but are not limited to: any *Schizochytrium* species, including *Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum*; any *Thraustochytrium* species (including former *Ulkenia* species such as *U. visurgensis, U. amoeboida, U. sarkariana, U. profunda, U. radiata, U. minuta* and *Ulkenia sp.* BP-5601), and including *Thraustochytrium striatum, Thraustochytrium aureum, Thraustochytrium roseum*; and any *Japonochytrium* species. Particularly preferred strains of Thraustochytriales include, but are not limited to: *Schizochytrium sp.* (S31)(ATCC 20888); *Schizochytrium sp.* (S8)(ATCC 20889); *Schizochytrium sp.* (LC-RM)(ATCC 18915); *Schizochytrium sp.* (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky)(ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi)(IFO 32693); *Thraustochytrium sp.* (23B)(ATCC 20891); *Thraustochytrium striatum* (Schneider)(ATCC 24473); *Thraustochytrium aureum* (Goldstein)(ATCC 34304); *Thraustochytrium roseum* (Goldstein)(ATCC 28210); and *Japonochytrium sp.* (L1)(ATCC 28207). Other examples of suitable host microorganisms for genetic modification include, but are not limited to, yeast including *Saccharomyces cerevisiae, Saccharomyces carlsbergensis*, or other yeast such as *Candida, Kluyveromyces*, or other fungi, for example, filamentous fungi such as *Aspergillus, Neurospora, Penicillium*, etc. Bacterial cells also may be used as hosts. These include, but are not limited to, *Escherichia coli*, which can be useful in fermentation processes. Alternatively, and only by way of example, a host such as a *Lactobacillus* species or *Bacillus* species can be used as a host.

Another embodiment of the present invention relates to a genetically modified plant, wherein the plant has been genetically modified to recombinantly express a PKS system comprising at least one biologically active domain of a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system. The domain of the PUFA PKS system can include any of the domains, including homologues thereof, for PUFA PKS systems as described above (e.g., for *Schizochytrium* and/or *Thraustochytrium*), and can also include any domain of a PUFA PKS system from any non-bacterial microorganism (including any eukaryotic microorganism and any other Thraustochytrid microorganism) or any domain of a PUFA PKS system from a microorganism identified by a screening method as described in U.S. Patent Application Serial No. 10/124,800, *supra*. The plant can also be further modified with at least one domain or biologically active fragment thereof of another PKS system, including, but not limited to, bacterial PUFA PKS or PKS systems, Type I PKS systems, Type II PKS systems, modular PKS systems, and/or any non-bacterial PUFA PKS system (e.g., eukaryotic, Thraustochytrid, Thraustochytriaceae or Labyrinthulaceae, *Schizochytrium*, etc.).

As used herein, a genetically modified plant can include any genetically modified plant including higher plants and particularly, any consumable plants or plants useful for producing a desired bioactive molecule of the present invention. Such a genetically modified plant has a genome which is modified (i.e., mutated or changed) from its normal (i.e., wild-type or naturally occurring) form such that the desired result is achieved (i.e., increased or modified PUFA PKS activity and/or production of a desired product using the PKS system). Genetic modification of a plant can be accomplished using classical strain development and/or molecular genetic techniques. Methods for producing a transgenic plant, wherein a recombinant nucleic acid molecule encoding a desired amino acid sequence is incorporated into the genome of the plant, are known in the art. A preferred plant to genetically modify according to the present invention is preferably a plant suitable for consumption by animals, including humans.

Preferred plants to genetically modify according to the present invention (i.e., plant host cells) include, but are not limited to any higher plants, and particularly consumable

plants, including crop plants and especially plants used for their oils. Such plants can include, for example: canola, soybeans, rapeseed, linseed, corn, safflowers, sunflowers and tobacco. Other preferred plants include those plants that are known to produce compounds used as pharmaceutical agents, flavoring agents, neutraceutical agents, functional food ingredients or cosmetically active agents or plants that are genetically engineered to produce these compounds/agents.

According to the present invention, a genetically modified microorganism or plant includes a microorganism or plant that has been modified using recombinant technology or by classical mutagenesis and screening techniques. As used herein, genetic modifications which result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e., the protein encoded by the gene) can be referred to as inactivation (complete or partial), deletion, interruption, blockage or down-regulation of a gene. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such gene, can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein does not exist), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no enzymatic activity or action). Genetic modifications that result in an increase in gene expression or function can be referred to as amplification, overproduction, overexpression, activation, enhancement, addition, or up-regulation of a gene.

The genetic modification of a microorganism or plant according to the present invention preferably affects the activity of the PKS system expressed by the microorganism or plant, whether the PKS system is endogenous and genetically modified, endogenous with the introduction of recombinant nucleic acid molecules into the organism (with the option of modifying the endogenous system or not), or provided completely by recombinant technology. To alter the PUFA production profile of a PUFA PKS system or organism expressing such system includes causing any detectable or measurable change in the production of any one or more PUFAs by the host microorganism or plant as compared to

in the absence of the genetic modification (i.e., as compared to the unmodified, wild-type microorganism or plant or the microorganism or plant that is unmodified at least with respect to PUFA synthesis - i.e., the organism might have other modifications not related to PUFA synthesis). To affect the activity of a PKS system includes any genetic modification that causes any detectable or measurable change or modification in the PKS system expressed by the organism as compared to in the absence of the genetic modification. A detectable change or modification in the PKS system can include, but is not limited to: a change or modification (introduction of, increase or decrease) of the expression and/or biological activity of any one or more of the domains in a modified PUFA PKS system as compared to the endogenous PUFA PKS system in the absence of genetic modification, the introduction of PKS system activity into an organism such that the organism now has measurable/detectable PKS system activity (i.e., the organism did not contain a PKS system prior to the genetic modification), the introduction into the organism of a functional domain from a different PKS system than a PKS system endogenously expressed by the organism such that the PKS system activity is modified (e.g., a bacterial PUFA PKS domain or a type I PKS domain is introduced into an organism that endogenously expresses a non-bacterial PUFA PKS system), a change in the amount of a bioactive molecule (e.g., a PUFA) produced by the PKS system (e.g., the system produces more (increased amount) or less (decreased amount) of a given product as compared to in the absence of the genetic modification), a change in the type of a bioactive molecule (e.g., a change in the type of PUFA) produced by the PKS system (e.g., the system produces an additional or different PUFA, a new or different product, or a variant of a PUFA or other product that is naturally produced by the system), and/or a change in the ratio of multiple bioactive molecules produced by the PKS system (e.g., the system produces a different ratio of one PUFA to another PUFA, produces a completely different lipid profile as compared to in the absence of the genetic modification, or places various PUFAs in different positions in a triacylglycerol as compared to the natural configuration). Such a genetic modification includes any type of genetic modification and specifically includes modifications made by recombinant technology and by classical mutagenesis.

It should be noted that reference to increasing the activity of a functional domain or protein in a PUFA PKS system refers to any genetic modification in the organism containing the domain or protein (or into which the domain or protein is to be introduced) which results in increased functionality of the domain or protein system and can include higher activity of the domain or protein (e.g., specific activity or *in vivo* enzymatic activity), reduced inhibition or degradation of the domain or protein system, and overexpression of the domain or protein. For example, gene copy number can be increased, expression levels can be increased by use of a promoter that gives higher levels of expression than that of the native promoter, or a gene can be altered by genetic engineering or classical mutagenesis to increase the activity of the domain or protein encoded by the gene.

Similarly, reference to decreasing the activity of a functional domain or protein in a PUFA PKS system refers to any genetic modification in the organism containing such domain or protein (or into which the domain or protein is to be introduced) which results in decreased functionality of the domain or protein and includes decreased activity of the domain or protein, increased inhibition or degradation of the domain or protein and a reduction or elimination of expression of the domain or protein. For example, the action of domain or protein of the present invention can be decreased by blocking or reducing the production of the domain or protein, "knocking out" the gene or portion thereof encoding the domain or protein, reducing domain or protein activity, or inhibiting the activity of the domain or protein. Blocking or reducing the production of a domain or protein can include placing the gene encoding the domain or protein under the control of a promoter that requires the presence of an inducing compound in the growth medium. By establishing conditions such that the inducer becomes depleted from the medium, the expression of the gene encoding the domain or protein (and therefore, of protein synthesis) could be turned off. The present inventors demonstrate the ability to delete (knock out) targeted genes in a Thraustochytrid microorganism in the Examples section. Blocking or reducing the activity of domain or protein could also include using an excision technology approach similar to that described in U.S. Patent No. 4,743,546, incorporated herein by reference. To use this approach, the gene encoding the protein of interest is cloned between specific genetic

sequences that allow specific, controlled excision of the gene from the genome. Excision could be prompted by, for example, a shift in the cultivation temperature of the culture, as in U.S. Patent No. 4,743,546, or by some other physical or nutritional signal.

In one embodiment of the present invention, a genetic modification includes a modification of a nucleic acid sequence encoding an amino acid sequence that has a biological activity of at least one domain of a non-bacterial PUFA PKS system as described herein (e.g., a domain, more than one domain, a protein, or the entire PUFA PKS system, of an endogenous PUFA PKS system of a Thraustochytrid host). Such a modification can be made to an amino acid sequence within an endogenously (naturally) expressed non-bacterial PUFA PKS system, whereby a microorganism that naturally contains such a system is genetically modified by, for example, classical mutagenesis and selection techniques and/or molecular genetic techniques, include genetic engineering techniques. Genetic engineering techniques can include, for example, using a targeting recombinant vector to delete a portion of an endogenous gene (demonstrated in the Examples), or to replace a portion of an endogenous gene with a heterologous sequence (demonstrated in the Examples). Examples of heterologous sequences that could be introduced into a host genome include sequences encoding at least one functional domain from another PKS system, such as a different non-bacterial PUFA PKS system (e.g., from a eukaryote, including another Thraustochytrid), a bacterial PUFA PKS system, a type I PKS system, a type II PKS system, or a modular PKS system. A heterologous sequence can also include an entire PUFA PKS system (e.g., all genes associated with the PUFA PKS system) that is used to replace the entire endogenous PUFA PKS system (e.g., all genes of the endogenous PUFA PKS system) in a host. A heterologous sequence can also include a sequence encoding a modified functional domain (a homologue) of a natural domain from a PUFA PKS system of a host Thraustochytrid (e.g., a nucleic acid sequence encoding a modified domain from OrfB of a *Schizochytrium*, wherein the modified domain will, when used to replace the naturally occurring domain expressed in the *Schizochytrium*, alter the PUFA production profile by the *Schizochytrium*). Other heterologous sequences to introduce into the genome of a host includes a sequence encoding a protein or functional domain that is not a domain of a PKS system, but which will

affect the activity of the endogenous PKS system. For example, one could introduce into the host genome a nucleic acid molecule encoding a phosphopantetheinyl transferase (discussed below). Specific modifications that could be made to an endogenous PUFA PKS system are discussed in detail herein.

In another aspect of this embodiment of the invention, the genetic modification can include: (1) the introduction of a recombinant nucleic acid molecule encoding an amino acid sequence having a biological activity of at least one domain of a PUFA PKS system; and/or (2) the introduction of a recombinant nucleic acid molecule encoding a protein or functional domain that affects the activity of a PUFA PKS system, into a host. The host can include: (1) a host cell that does not express any PKS system, wherein all functional domains of a PKS system are introduced into the host cell, and wherein at least one functional domain is from a non-bacterial PUFA PKS system; (2) a host cell that expresses a PKS system (endogenous or recombinant) having at least one functional domain of a non-bacterial PUFA PKS system, wherein the introduced recombinant nucleic acid molecule can encode at least one additional non-bacterial PUFA PKS domain function or another protein or domain that affects the activity of the host PKS system; and (3) a host cell that expresses a PKS system (endogenous or recombinant) which does not necessarily include a domain function from a non-bacterial PUFA PKS, and wherein the introduced recombinant nucleic acid molecule includes a nucleic acid sequence encoding at least one functional domain of a non-bacterial PUFA PKS system. In other words, the present invention intends to encompass any genetically modified organism (e.g., microorganism or plant), wherein the organism comprises at least one non-bacterial PUFA PKS domain function (either endogenously or introduced by recombinant modification), and wherein the genetic modification has a measurable effect on the non-bacterial PUFA PKS domain function or on the PKS system when the organism comprises a functional PKS system.

The present invention encompasses many possible non-bacterial and bacterial microorganisms as either possible host cells for the PUFA PKS systems described herein and/or as sources for additional genetic material encoding PUFA PKS system proteins and domains for use in the genetic modifications and methods described herein. For example,

microbial organisms with a PUFA PKS system similar to that found in *Schizochytrium*, such as the *Thraustochytrium* microorganism discovered by the present inventors and described in Example 1, can be readily identified/isolated/screened by methods to identify other non-bacterial microorganisms that have a polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system that are described in detail in U.S. Patent Application Publication No. 20020194641, *supra* (corresponding to U.S. Patent Application Serial No. 10/124,800).

Locations for collection of the preferred types of microbes for screening for a PUFA PKS system according to the present invention include any of the following: low oxygen environments (or locations near these types of low oxygen environments including in the guts of animals including invertebrates that consume microbes or microbe-containing foods (including types of filter feeding organisms), low or non-oxygen containing aquatic habitats (including freshwater, saline and marine), and especially at-or near-low oxygen environments (regions) in the oceans. The microbial strains would preferably not be obligate anaerobes but be adapted to live in both aerobic and low or anoxic environments. Soil environments containing both aerobic and low oxygen or anoxic environments would also excellent environments to find these organisms in and especially in these types of soil in aquatic habitats or temporary aquatic habitats.

A particularly preferred non-bacterial microbial strain to screen for use as a host and/or a source of PUFA PKS genes according to the present invention would be a strain (selected from the group consisting of algae, fungi (including yeast), protozoa or protists) that, during a portion of its life cycle, is capable of consuming whole bacterial cells (bacterivory) by mechanisms such as phagocytosis, phagotrophic or endocytic capability and/or has a stage of its life cycle in which it exists as an amoeboid stage or naked protoplast. This method of nutrition would greatly increase the potential for transfer of a bacterial PKS system into a eukaryotic cell if a mistake occurred and the bacterial cell (or its DNA) did not get digested and instead are functionally incorporated into the eukaryotic cell.

Included in the present invention as sources of PUFA PKS genes (and proteins and domains encoded thereby) are any Thraustochytrids other than those specifically described herein that contain a PUFA PKS system. Such Thraustochytrids include, but are not limited

to, but are not limited to, any microorganism of the order Thraustochytriales, including any microorganism in the families Thraustochytriaceae and Labyrinthulaceae, which further comprise a genus including, but not limited to: *Thraustochytrium, , Japonochytrium, Aplanochytrium, Elina* and *Schizochytrium* within the Thraustochytriaceae and *Labyrinthula, Labyrinthuloides*, and *Labyrinthomyxa* within the Labyrinthulaceae. Preferred species within these genera include, but are not limited to: any species within *Labyrinthula*, including *Labrinthula* sp., *Labyrinthula algeriensis, Labyrinthula cienkowskii, Labyrinthula chattonii, Labyrinthula coenocystis, Labyrinthula macrocystis, Labyrinthula macrocystis atlantica, Labyrinthula macrocystis macrocystis, Labyrinthula magnifica, Labyrinthula minuta, Labyrinthula roscoffensis, Labyrinthula valkanovii, Labyrinthula vitellina, Labyrinthula vitellina pacifica, Labyrinthula vitellina vitellina, Labyrinthula zopfii*; any *Labyrinthuloides* species, including *Labyrinthuloides* sp., *Labyrinthuloides minuta, Labyrinthuloides schizochytrops*; any *Labyrinthomyxa* species, including *Labyrinthomyxa* sp., *Labyrinthomyxa pohlia, Labyrinthomyxa sauvageaui*, any *Aplanochytrium* species, including *Aplanochytrium* sp. and *Aplanochytrium kerguelensis*; any *Elina* species, including *Elina* sp., *Elina marisalba, Elina sinorifica*; any *Japanochytrium* species, including *Japanochytrium* sp., *Japanochytrium marinum*; any *Schizochytrium* species, including *Schizochytrium* sp., *Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum, Schizochytrium octosporum;* and any *Thraustochytrium* species, including *Thraustochytrium* sp., *Thraustochytrium aggregatum, Thraustochytrium arudimentale, Thraustochytrium aureum, Thraustochytrium benthicola, Thraustochytrium globosum, Thraustochytrium kinnei, Thraustochytrium motivum, Thraustochytrium pachydermum, Thraustochytrium proliferum, Thraustochytrium roseum, Thraustochytrium striatum, Ulkenia* sp., *Ulkenia minuta, Ulkenia profunda, Ulkenia radiate, Ulkenia sarkariana,* and *Ulkenia visurgensis.*

It is noted that, without being bound by theory, the present inventors consider *Labyrinthula* and other Labyrinthulaceae as sources of PUFA PKS genes because the Labyrinthulaceae are closely related to the Thraustochytriaceae which are known to possess PUFA PKS genes, the Labyrinthulaceae are known to be bactivorous/phagocytotic, and some

members of the Labyrinthulaceae have fatty acid/PUFA profiles consistent with having a PUFA PKS system.

Strains of microbes (other than the members of the Thraustochytrids) capable of bacterivory (especially by phagocytosis or endocytosis) can be found in the following microbial classes (including but not limited to example genera):

In the algae and algae-like microbes (including Stramenopiles): of the class Euglenophyceae (for example genera *Euglena*, and *Peranema*), the class Chrysophyceae (for example the genus *Ochromonas*), the class Dinobryaceae (for example the genera *Dinobryon*, *Platychrysis*, and *Chrysochromulina*), the Dinophyceae (including the genera *Crypthecodinium*, *Gymnodinium*, *Peridinium*, *Ceratium*, *Gyrodinium*, and *Oxyrrhis*), the class Cryptophyceae (for example the genera *Cryptomonas*, and *Rhodomonas*), the class Xanthophyceae (for example the genus *Olisthodiscus*) (and including forms of algae in which an amoeboid stage occurs as in the flagellates Rhizochloridaceae, and zoospores/gametes of *Aphanochaete pascheri*, *Bumilleria stigeoclonium* and *Vaucheria geminata*), the class Eustigmatophyceae, and the class Prymnesiopyceae (including the genera Prymnesium and Diacronema).

In the Stramenopiles including the: Proteromonads, Opalines, Developayella, Diplophorys, Labyrinthulids, Thraustochytrids, Bicosecids, Oomycetes, Hypochytridiomycetes, Commation, Reticulosphaera, Pelagomonas, Pelapococcus, Ollicola, Aureococcus, Parmales, Raphidiophytes, Synurids, Rhizochromulinaales, Pedinellales, Dictyochales, Chrysomeridales, Sarcinochrysidales, Hydrurales, Hibberdiales, and Chromulinales.

In the Fungi: Class Myxomycetes (form myxamoebae) -- slime molds, class Acrasieae including the orders Acrasiceae (for example the genus Sappinia), class Guttulinaceae (for example the genera *Guttulinopsis, and Guttulina*), class Dictysteliaceae (for example the genera *Acrasis, Dictyostelium, Polysphondylium, and Coenonia*), and class Phycomyceae including the orders Chytridiales, Ancylistales, Blastocladiales, Monoblepharidales, Saprolegniales, Peronosporales, Mucorales, and Entomophthorales.

In the Protozoa: Protozoa strains with life stages capable of bacterivory (including by phageocytosis) can be selected from the types classified as ciliates, flagellates or amoebae. Protozoan ciliates include the groups: Chonotrichs, Colpodids, Cyrtophores, Haptorids, Karyorelicts, Oligohymenophora, Polyhymenophora (spirotrichs), Prostomes and Suctoria. Protozoan flagellates include the Biosoecids, Bodonids, Cercomonads, Chrysophytes (for example the genera Anthophysa, Chrysamoemba, Chrysosphaerella, Dendromonas, Dinobryon, Mallomonas, Ochromonas, Paraphysomonas, Poterioochromonas, Spumella, Syncrypta, Synura, and Uroglena), Collar flagellates, Cryptophytes (for example the genera Chilomonas, Cryptomonas, Cyanomonas, and Goniomonas), Dinoflagellates, Diplomonads, Euglenoids, Heterolobosea, Pedinellids, Pelobionts, Phalansteriids, Pseudodendromonads, Spongomonads and Volvocales (and other flagellates including the unassigned flagellate genera of Artodiscus, Clautriavia, Helkesimastix, Kathablepharis and Multicilia). Amoeboid protozoans include the groups: Actinophryids, Centrohelids, Desmothoricids, Diplophryids, Eumamoebae, Heterolobosea, Leptomyxids, Nucleariid filose amoebae, Pelebionts, Testate amoebae and Vampyrellids (and including the unassigned amoebid genera Gymnophrys, Biomyxa, Microcometes, Reticulomyxa, Belonocystis, Elaeorhanis, Allelogromia, Gromia or Lieberkuhnia). The protozoan orders include the following: Percolomonadeae, Heterolobosea, Lyromonadea, Pseudociliata, Trichomonadea, Hypermastigea, Heteromiteae, Telonemea, Cyathobodonea, Ebridea, Pyytomyxea, Opalinea, Kinetomonadea, Hemimastigea, ,Protostelea, Myxagastrea, Dictyostelea, Choanomonadea, Apicomonadea, Eogregarinea, Neogregarinea, Coelotrolphea, Eucoccidea, Haemosporea, Piroplasmea, Spirotrichea, Prostomatea, Litostomatea, Phyllopharyngea, Nassophorea, Oligohymenophorea, Colpodea, Karyorelicta, Nucleohelea, Centrohelea, Acantharea, Sticholonchea, Polycystinea, Phaeodarea, Lobosea, Filosea, Athalamea, Monothalamea, Polythalamea, Xenophyophorea, Schizocladea, Holosea, Entamoebea, Myxosporea, Actinomyxea, Halosporea, Paramyxea, Rhombozoa and Orthonectea.

A preferred embodiment of the present invention includes strains of the microorganisms listed above that have been collected from one of the preferred habitats listed above.

In some embodiments of this method of the present invention, PUFA PKS systems from bacteria, including genes and portions thereof (encoding entire PUFA PKS systems, proteins thereof and/or domains thereof) can be used to genetically modify other PUFA PKS systems (e.g., any non-bacterial PUFA PKS system) and/or microorganisms containing the same (or vice versa) in the embodiments of the invention. In one aspect, novel PUFA PKS systems can be identified in bacteria that are expected to be particularly useful for creating genetically modified microorganisms (e.g., genetically modified Thraustochytrids) and/or novel hybrid constructs encoding PUFA PKS systems for use in the methods and genetically modified microorganisms and plants of the present invention. In one aspect, bacteria that may be particularly useful in the embodiments of the present invention have PUFA PKS systems, wherein the PUFA PKS system is capable of producing PUFAs at temperatures exceeding about 20°C, preferably exceeding about 25°C and even more preferably exceeding about 30°C. As described previously herein, the marine bacteria, *Shewanella* and *Vibrio marinus*, described in U.S. Patent No. 6,140,486, do not produce PUFAs at higher temperatures, which limits the usefulness of PUFA PKS systems derived from these bacteria, particularly in plant applications under field conditions. Therefore, in one embodiment, the screening method of the present invention can be used to identify bacteria that have a PUFA PKS system, wherein the bacteria are capable of growth and PUFA production at higher temperatures (e.g., above about 15°C, 20°C, 25°C, or 30°C or even higher). However, even if the bacteria sources do not grow well and/or produce PUFAs at the higher temperatures, the present invention encompasses the identification, isolation and use of the PUFA PKS systems (genes and proteins/domains encoded thereby), wherein the *PUFA PKS systems* from the bacteria have enzymatic/biological activity at temperatures above about 15°C, 20°C, 25°C, or 30°C or even higher. In one aspect of this embodiment, inhibitors of eukaryotic growth such as nystatin (antifungal) or cycloheximide (inhibitor of eukaryotic protein synthesis) can be added to agar plates used to culture/select initial strains from water samples/soil samples collected from the types of habitats/niches such as marine or estuarian habits, or any other habitat where such bacteria can be found. This process would help select for enrichment of bacterial strains without (or minimal) contamination of eukaryotic strains.

This selection process, in combination with culturing the plates at elevated temperatures (e.g. 30°C), and then selecting strains that produce at least one PUFA would initially identify candidate bacterial strains with a PUFA PKS system that is operative at elevated temperatures (as opposed to those bacterial strains in the prior art which only exhibit PUFA production at temperatures less than about 20°C and more preferably below about 5°C).

However, even in bacteria that do not grow well (or at all) at higher temperatures, or that do not produce at least one PUFA at higher temperatures, such strains can be identified and selected as comprising a PUFA PKS system by the identification of the ability of the bacterium to produce PUFAs under any conditions and/or by screening the genome of the bacterium for genes that are homologous to other known PUFA PKS genes from bacteria or non-bacterial organisms (e.g., see Example 7). To evaluate PUFA PKS function at higher temperatures for genes from any bacterial source, one can produce cell-free extracts and test for PUFA production at various temperatures, followed by selection of microorganisms that contain PUFA PKS genes that have enzymatic/biological activity at higher temperature ranges (e.g., 15°C, 20°C, 25°C, or 30°C or even higher).

Suitable bacteria to use as hosts for genetic modification include any bacterial strain as discussed above. Particularly suitable bacteria to use as a *source* of PUFA PKS genes (and proteins and domains encoded thereby) for the production of genetically modified sequences and organisms according to the present invention include any bacterium that comprises a PUFA PKS system. Such bacteria are typically isolated from marine or estuarian habitats and can be readily identified by their ability to product PUFAs and/or by the presence of one or more genes having homology to known PUFA PKS genes in the organism. Such bacteria can include, but are not limited to, bacteria of the genera *Shewanella* and *Vibrio*. Preferred bacteria for use in the present invention include those with PUFA PKS systems that are biologically active at higher temperatures (e.g., above about 15°C, 20°C, 25°C, or 30°C or even higher). The present inventors have identified two exemplary bacteria (e.g. *Shewanella olleyana* and *Shewanella japonica*; see Examples 7 and 8) that will be particularly suitable for use as sources of PUFA PKS genes, and others can

be readily identified or are known to comprise PUFA PKS genes and may be useful in an embodiment of the present invention (e.g., *Shewanella gelidimarina*).

Furthermore, it is recognized that not all bacterial or non-bacterial microorganisms can be readily cultured from natural habitats. However, genetic characteristics of such un-culturable microorganisms can be evaluated by isolating genes from DNA prepared *en mass* from mixed or crude environmental samples. Particularly suitable to the present invention, PUFA PKS genes derived from un-culturable microorganisms can be isolated from environmental DNA samples by degenerate PCR using primers designed to generally match regions of high similarity in known PUFA PKS genes (e.g., see Example 7). Alternatively, whole DNA fragments can be cloned directly from purified environmental DNA by any of several methods known to the art. Sequence of the DNA fragments thus obtained can reveal homologs to known genes such as PUFA PKS genes. Homologs of OrfB and OrfC (referring to the domain structure of *Schizochytrium* and *Thraustochytrium*, for example) may be particularly useful in defining the PUFA PKS end product. Whole coding regions of PUFA PKS genes can then be expressed in host organisms (such as *Escherichia coli* or yeast) in combination with each other or with known PUFA PKS gene or gene fragment combinations to evaluate their effect on PUFA production. As described above, activity in cell-free extracts can be used to determine function at desired temperatures. Isolated PUFA PKS genes can also be transformed directly into appropriate *Schizochytrium* or other suitable strains to measure function. PUFA PKS system-encoding constructs identified or produced in such a manner, including hybrid constructs, can also be used to transform other organisms, such as plants.

Therefore, using the non-bacterial PUFA PKS systems of the present invention, which, for example, makes use of genes from Thraustochytrid PUFA PKS systems, as well as PUFA PKS systems and PKS systems from bacteria, gene mixing can be used to extend the range of PUFA products to include EPA, DHA, ARA, GLA, SDA and others (described in detail below), as well as to produce a wide variety of bioactive molecules, including antibiotics, other pharmaceutical compounds, and other desirable products. The method to obtain these bioactive molecules includes not only the mixing of genes from various

organisms but also various methods of genetically modifying the non-bacterial PUFA PKS genes disclosed herein. Knowledge of the genetic basis and domain structure of the non-bacterial PUFA PKS system of the present invention provides a basis for designing novel genetically modified organisms which produce a variety of bioactive molecules. Although mixing and modification of any PKS domains and related genes are contemplated by the present inventors, by way of example, various possible manipulations of the PUFA-PKS system are discussed below with regard to genetic modification and bioactive molecule production.

Accordingly, encompassed by the present invention are methods to genetically modify microbial or plant cells by: genetically modifying at least one nucleic acid sequence in the organism that encodes an amino acid sequence having the biological activity of at least one functional domain of a non-bacterial PUFA PKS system according to the present invention, and/or expressing at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding such amino acid sequence. Various embodiments of such sequences, methods to genetically modify an organism, and specific modifications have been described in detail above. Typically, the method is used to produce a particular genetically modified organism that produces a particular bioactive molecule or molecules.

One embodiment of the present invention relates to a genetically modified Thraustochytrid microorganism, wherein the microorganism has an endogenous polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) system, and wherein the endogenous PUFA PKS system has been genetically modified to alter the expression profile of a polyunsaturated fatty acid (PUFA) by the microorganism as compared to the Thraustochytrid microorganism in the absence of the modification. Thraustochytrid microorganisms useful as host organisms in the present invention endogenously contain and express a PUFA PKS system. The genetic modification can be a genetic modification of one or more of the functional domains of the endogenous PUFA PKS system, whereby the modification alters the PUFA production profile of the endogenous PUFA PKS system. In addition, or as an alternative, the genetic modification can be an introduction of at least one exogenous nucleic acid sequence (e.g., a recombinant nucleic acid molecule) to the

microorganism, wherein the exogenous nucleic acid sequence encodes at least one biologically active domain or protein from a second PKS system and/or a protein that affects the activity of the PUFA PKS system (e.g., a phosphopantetheinyl transferases (PPTase)). The second PKS system can be any PKS system, including other PUFA PKS systems and including homologues of genes from the Thraustochytrid PUFA PKS system to be genetically modified.

This embodiment of the invention is particularly useful for the production of commercially valuable lipids enriched in a desired PUFA, such as EPA, via the present inventors' development of genetically modified microorganisms and methods for efficiently producing lipids (triacylglyerols (TAG) as well as membrane-associated phospholipids (PL)) enriched in PUFAs.

This particular embodiment of the present invention is derived in part from the following knowledge: (1) utilization of the inherent TAG production capabilities of selected microorganisms, and particularly, of Thraustochytrids, such as the commercially developed *Schizochytrium* strain described herein; (2) the present inventors' detailed understanding of PUFA PKS biosynthetic pathways (i.e., PUFA PKS systems) in eukaryotes and in particular, in members of the order Thraustochytriales; and, (3) utilization of a homologous genetic recombination system in *Schizochytrium*. Based on the inventors' knowledge of the systems involved, the same general approach may be exploited to produce PUFAs other than EPA.

In one embodiment of the invention, the endogenous Thraustochytrid PUFA PKS genes, such as the *Schizochytrium* genes encoding PUFA PKS enzymes that normally produce DHA and DPA, are modified by random or targeted mutagenesis, replaced with genes from other organisms that encode homologous PKS proteins (e.g., from bacteria or other sources), or replaced with genetically modified *Schizochytrium*, *Thraustochytrium* or other Thraustochytrid PUFA PKS genes. The product of the enzymes encoded by these introduced and/or modified genes can be EPA, for example, or it could be some other related molecule, including other PUFAs. One feature of this method is the utilization of endogenous components of Thraustochytrid PUFA synthesis and accumulation machinery that is essential for efficient production and incorporation of the PUFA into PL and TAG.

In particular, this embodiment of the invention is directed to the modification of the type of PUFA produced by the organism, while retaining the high oil productivity of the parent strain.

Although some of the following discussion uses the organism *Schizochytrium* as an exemplary host organism, any Thraustochytrid can be modified according to the present invention, including members of the genera *Thraustochytrium*, *Labyrinthuloides*, and *Japonochytrium*. For example, the genes encoding the PUFA PKS system for a species of *Thraustochytrium* have been identified (see Example 6), and this organism can also serve as a host organism for genetic modification using the methods described herein, although it is more likely that the *Thraustochytrium* PKS genes will be used to modify the endogenous PUFA PKS genes of another Thraustochytrid, such as *Schizochytrium*. Furthermore, using methods for screening organisms as set forth in U.S. Application Serial No. 10/124,800, *supra*, one can identify other organisms useful in the present method and all such organisms are encompassed herein.

This embodiment of the present invention can be illustrated as follows. By way of example, based on the present inventors' current understanding of PUFA synthesis and accumulation in *Schizochytrium*, the overall biochemical process can be divided into three parts.

First, the PUFAs that accumulate in *Schizochytrium* oil (DHA and DPA) are the product of a PUFA PKS system as discussed above. The PUFA PKS system in *Schizochytrium* converts malonyl-CoA into the end product PUFA without release of significant amounts of intermediate compounds. In *Schizochytrium*, three genes have been identified (Orfs A, B and C; also represented by SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, respectively) that encode all of the enzymatic domains known to be required for actual synthesis of PUFAs. Similar sets of genes (encoding proteins containing homologous sets of enzymatic domains) have been cloned and characterized from several other non-eukaryotic organisms that produce PUFAs, namely, several strains of marine bacteria. In addition, the present inventors have identified and now sequenced PUFA PKS genes in at least one other marine protist (Thraustochytrium strain 23B) (described in detail below).

The PUFA products of marine bacteria include EPA (e.g., produced by *Shewanella* SRC2738 and *Photobacter profundum*) as well as DHA (*Vibrio marinus*, now known as *Moritella marina*) (described in U.S. Patent No. 6,140,486, *supra*; and in U.S. Patent No. 6,566,583, *supra*). It is an embodiment of the invention that any PUFA PKS gene set could be envisioned to substitute for the *Schizochytrium* genes described in the example herein, as long as the physiological growth requirements of the production organism (e.g., *Schizochytrium*) in fermentation conditions were satisfied. In particular, the PUFA-producing bacterial strains described above grow only at relatively low temperatures (typically less than 20°C) which further indicates that their PUFA PKS gene products will not function at standard growth temperatures for *Schizochytrium* (25-30°C). However, the inventors have recently identified at least two other marine bacteria that grow and produce EPA at standard growth temperatures for *Schizochytrium* and other Thraustochytrids (see Example 7). These alternate marine bacteria have been shown to possess PUFA-PKS-like genes that will serve as material for modification of *Schizochytrium* and other Thraustochytrids by methods described herein. It will be apparent to those skilled in the art from this disclosure that other currently unstudied or unidentified PUFA-producing bacteria could also contain PUFA PKS genes useful for modification of Thraustochytrids.

Second, in addition to the genes that encode the enzymes directly involved in PUFA synthesis, an "accessory" enzyme is required. The gene encodes a phosphopantetheine transferase (PPTase) that activates the acyl-carrier protein (ACP) domains present in the PUFA PKS complex. Activation of the ACP domains by addition of this co-factor is required for the PUFA PKS enzyme complex to function. All of the ACP domains of the PUFA PKS systems identified so far show a high degree of amino acid sequence conservation and, without being bound by theory, the present inventors believe that the PPTase of *Schizochytrium* and other Thraustochytrids will recognize and activate ACP domains from other PUFA PKS systems. As proof of principle that heterologous PPTases and PUFA PKS genes can function together to produce a PUFA product, the present inventors demonstrate herein the use of two different heterologous PPTases with the PUFA PKS genes from *Schizochytrium* to produce a PUFA in a bacterial host cell.

Third, in *Schizochytrium*, the products of the PUFA PKS system are efficiently channeled into both the phospholipids (PL) and triacylglycerols (TAG). The present inventors' data suggest that the PUFA is transferred from the ACP domains of the PKS complex to coenzyme A (CoA). As in other eukaryotic organisms, this acyl-CoA would then serve as the substrate for the various acyl-transferases that form the PL and TAG molecules. In contrast, the data indicate that in bacteria, transfer to CoA does not occur; rather, there is a direct transfer from the ACP domains of the PKS complex to the acyl-transferases that form PL. The enzymatic system in *Schizochytrium* that transfers PUFA from ACP to CoA clearly can recognize both DHA and DPA and therefore, the present inventors believe that it is predictable that any PUFA product of the PUFA PKS system (as attached to the PUFA PKS ACP domains) will serve as a substrate.

Therefore, in one embodiment of the present invention, the present inventors propose to alter the genes encoding the components of the PUFA PKS enzyme complex (part 1) while utilizing the endogenous PPTase from *Schizochytrium* or another Thraustochytrid host (part 2) and PUFA-ACP to PUFA-CoA transferase activity and TAG / PL synthesis systems (or other endogenous PUFA ACP to TAG/PL mechanism) (part 3). These methods of the present invention are supported by experimental data, some of which are presented in the Examples section in detail.

First, the present inventors have found that the PUFA PKS system can be transferred between organisms, and that some parts are interchangeable. More particularly, it has been previously shown that the PUFA PKS pathways of the marine bacteria, *Shewanella* SCR2738 (Yazawa, 1996, *Lipids* 31:S297-300) and *Vibrio marinus* (along with the PPTase from *Shewanella*) (U.S. Patent No. 6,140,486), can be successfully transferred to a heterologous host (i.e., to *E. coli*). Additionally, the degree of structural homology between the subunits of the PUFA PKS enzymes from these two organisms (*Shewanella* SCRC2738 and *Vibrio marinus*) is such that it has been possible to mix and match genes from the two systems (U.S. Patent No. 6,140,486, *supra*). The PUFA end product of the mixed sets of genes varied depending on the origins of the specific gene homologues. At least one open reading frame (*Shewanella's* Orf 7 and its *Vibrio marinus* homologue; see Fig. 13 of U.S. Patent No.

6,140,486; note that the nomenclature for this Orf has changed; it is labeled as Orf 8 in the patent, but was submitted to Genbank as Orf 7, and is now referred to by its GenBank designation) could be associated with determination of whether DHA or EPA would be the product of the composite system. The functional domains of all of the PUFA PKS enzymes identified so far show sequence homology to one another. Similarly, these data indicated that PUFA PKS systems, including those from the marine bacteria, can be transferred to, and will function in, *Schizochytrium* and other Thraustochytrids.

The present inventors have now expressed the PUFA PKS genes (Orfs A, B and C) from *Schizochytrium* in an *E. coli* host and have demonstrated that the cells made DHA and DPA in about the same ratio as the endogenous production of these PUFAs in *Schizochytrium* (see Example 2). Therefore, it has been demonstrated that the recombinant *Schizochytrium* PUFA PKS genes encode a functional PUFA synthesis system. Additionally, all or portions of the *Thraustochytrium* 23B OrfA and OrfC genes have been shown to function in *Schizochytrium* (see Example 6).

Second, the present inventors have previously found that PPTases can activate heterologous PUFA PKS ACP domains. Production of DHA in *E. coli* transformed with the PUFA PKS genes from *Vibrio marinus* occurred only when an appropriate PPTase gene (in this case, from Shewanella SCRC2738) was also present (see U.S. Patent No. 6,140,486, *supra*). This demonstrated that the *Shewanella* PPTase was able to activate the *Vibrio* PUFA PKS ACP domains. Additionally, the present inventors have now demonstrated the activation (pantetheinylation) of ACP domains from *Schizochytrium* Orf A using a PPTase (sfp) from *Bacillus subtilus* (see Example 2). The present inventors have also demonstrated activation (pantetheinylation) of ACP domains from *Schizochytrium* Orf A by a PPTase called Het I from Nostoc (see Example 2). The HetI enzyme was additionally used as the PPTase in the experiments discussed above for the production of DHA and DPA in *E. coli* using the recombinant *Schizochytrium* PUFA PKS genes (Example 2).

Third, data indicate that DHA-CoA and DPA-CoA may be metabolic intermediates in the *Schizochytrium* TAG and PL synthesis pathway. Published biochemical data suggest that in bacteria, the newly synthesized PUFAs are transferred directly from the PUFA PKS

ACP domains to the phospholipid synthesis enzymes. In contrast, the present inventors' data indicate that in *Schizochytrium*, a eukaryotic organism, there may be an intermediate between the PUFA on the PUFA PKS ACP domains and the target TAG and PL molecules. The typical carrier of fatty acids in the eukaryotic cytoplasm is CoA. The inventors examined extracts of *Schizochytrium* cells and found significant levels of compounds that co-migrated during HPLC fractionation with authentic standards of DHA-CoA, DPA-CoA, 16:0-CoA and 18:1-CoA. The identity of the putative DHA-CoA and DPA-CoA peaks were confirmed using mass spectroscopy. In contrast, the inventors were not able to detect DHA-CoA in extracts of *Vibrio marinus*, again suggesting that a different mechanism exists in bacteria for transfer of the PUFA to its final target (e.g., direct transfer to PL). The data indicate a mechanism likely exists in *Schizochytrium* for transfer of the newly synthesized PUFA to CoA (probably via a direct transfer from the ACP to CoA). Both TAG and PL synthesis enzymes could then access this PUFA-CoA. The observation that both DHA and DPA CoA are produced suggests that the enzymatic transfer machinery may recognize a range of PUFAs.

Fourth, the present inventors have now created knockouts of Orf A, Orf B, and Orf C in *Schizochytrium* (see Example 3). The knockout strategy relies on the homologous recombination that has been demonstrated to occur in *Schizochytrium* (see U.S. Patent Application Serial No. 10/124,807, *supra*). Several strategies can be employed in the design of knockout constructs. The specific strategy used to inactivate these three genes utilized insertion of a Zeocin™ resistance gene coupled to a tubulin promoter (derived from pMON50000, see U.S. Patent Application Serial No. 10/124,807) into a cloned portion of the Orf. The new construct containing the interrupted coding region was then used for the transformation of wild type *Schizochytrium* cells via particle bombardment (see U.S. Patent Application Serial No. 10/124,807). Bombarded cells were spread on plates containing both Zeocin™ and a supply of PUFA (see below). Colonies that grew on these plates were then streaked onto Zeocin™ plates that were not supplemented with PUFAs. Those colonies that required PUFA supplementation for growth were candidates for having had the PUFA PKS Orf inactivated via homologous recombination. In all three cases, this presumption was

confirmed by rescuing the knockout by transforming the cells with a full-length genomic DNA clones of the respective *Schizochytrium* Orfs. Furthermore, in some cases, it was found that the Zeocin™ resistance gene had been removed (see Example 5), indicating that the introduced functional gene had integrated into the original site by double homologous recombination (i.e. deleting the resistance marker). One key to the success of this strategy was supplementation of the growth medium with PUFAs. In the present case, an effective means of supplementation was found to be sequestration of the PUFA by mixing with partially methylated beta-cyclodextrin prior to adding to the growth medium (see Example 5). Together, these experiments demonstrate the principle that one of skill in the art, given the guidance provided herein, can inactivate one or more of the PUFA PKS genes in a PUFA PKS-containing microorganism such as *Schizochytrium*, and create a PUFA auxotroph which can then be used for further genetic modification (e.g., by introducing other PKS genes) according to the present invention (e.g., to alter the fatty acid profile of the recombinant organism).

One important element of the genetic modification of the organisms of the present invention is the ability to directly transform a Thraustochytrid genome. In U.S. Application Serial No. 10/124,807, *supra*, transformation of *Schizochytrium* via single crossover homologous recombination and targeted gene replacement via double crossover homologous recombination were demonstrated. As discussed above, the present inventors have now used this technique for homologous recombination to inactivate Orf A, Orf B and OrfC of the PUFA-PKA system in *Schizochytrium*. The resulting mutants are dependent on supplementation of the media with PUFA. Several markers of transformation, promoter elements for high level expression of introduced genes and methods for delivery of exogenous genetic material have been developed and are available. Therefore, the tools are in place for knocking out endogenous PUFA PKS genes in Thraustochytrids and other eukaryotes having similar PUFA PKS systems and replacing them with genes from other organisms (or with modified *Schizochytrium* genes) as proposed above.

In one approach for production of EPA-rich TAG, the PUFA PKS system of *Schizochytrium* can be altered by the addition of heterologous genes encoding a PUFA PKS

system whose product is EPA. It is anticipated that the endogenous PPTase will activate the ACP domains of that heterologous PUFA PKS system. Additionally, it is anticipated that the EPA will be converted to EPA-CoA and will readily be incorporated into *Schizochytrium* TAG and PL membranes. In one modification of this approach, techniques can be used to modify the relevant domains of the endogenous *Schizochytrium* system (either by introduction of specific regions of heterologous genes or by mutagenesis of the *Schizochytrium* genes themselves) such that its end product is EPA rather than DHA and DPA. This is an exemplary approach, as this technology can be applied to the production of other PUFA end products and to any eukaryotic microorganism that comprises a PUFA PKS system and that has the ability to efficiently channel the products of the PUFA PKS system into both the phospholipids (PL) and triacylglycerols (TAG). In particular, the invention is applicable to any Thraustochytrid microorganism or any other eukaryote that has an endogenous PUFA PKS system, which is described in detail below by way of example. In addition, the invention is applicable to any suitable host organism, into which the modified genetic material for production of various PUFA profiles as described herein can be transformed. For example, in the Examples, the PUFA PKS system from *Schizochytrium* is transformed into an *E. coli*. Such a transformed organism could then be further modified to alter the PUFA production profile using the methods described herein.

The present invention can make use of genes and nucleic acid sequences which encode proteins or domains from PKS systems other than the PUFA PKS system described herein and in U.S. Patent Application Serial No. 10/124,800, and include genes and nucleic acid sequences from bacterial and non-bacterial PKS systems, including PKS systems of Type II, Type I and modular, described above. Organisms which express each of these types of PKS systems are known in the art and can serve as sources for nucleic acids useful in the genetic modification process of the present invention.

In a preferred embodiment, genes and nucleic acid sequences which encode proteins or domains from PKS systems other than the PUFA PKS system or from other PUFA PKS systems are isolated or derived from organisms which have preferred growth characteristics for production of PUFAs. In particular, it is desirable to be able to culture the genetically

modified Thraustochytrid microorganism at temperatures greater than about 15°C, greater than 20°C, greater than 25°C, greater than 30°C, greater than 35°C, greater than 40°C, or in one embodiment, at any temperature between about 20°C and 40°C. Therefore, PKS proteins or domains having functional enzymatic activity at these temperatures are preferred. For example, the present inventors describe herein the use of PKS genes from *Shewanella olleyana* or *Shewanella japonica*, which are marine bacteria that naturally produce EPA and grow at temperatures up to 30°C and 35°C, respectively (see Example 7). PKS proteins or domains from these organisms are examples of proteins and domains that can be mixed with Thraustochytrid PUFA PKS proteins and domains as described herein to produce a genetically modified organism that has a specifically designed or modified PUFA production profile.

In another preferred embodiment, the genes and nucleic acid sequences that encode proteins or domains from a PUFA PKS system that produces one fatty acid profile are used to modify another PUFA PKS system and thereby alter the fatty acid profile of the host. For example, *Thraustochytrium* 23B (ATCC 20892) is significantly different from *Schizochytrium* sp. (ATCC 20888) in its fatty acid profile. *Thraustochytrium* 23B can have DHA:DPA(n-6) ratios as high as 40:1 compared to only 2-3:1 in *Schizochytrium* (ATCC 20888). *Thraustochytrium* 23B can also have higher levels of C20:5(n-3). However, *Schizochytrium* (ATCC 20888) is an excellent oil producer as compared to *Thraustochytrium* 23B. *Schizochytrium* accumulates large quantities of triacylglycerols rich in DHA and docosapentaenoic acid (DPA; 22:5ω6); e.g., 30% DHA + DPA by dry weight. Therefore, the present inventors describe herein the modification of the *Schizochytrium* endogenous PUFA PKS system with *Thraustochytrium* 23B PUFA PKS genes to create a genetically modified *Schizochytrium* with a DHA:DPA profile more similar to *Thraustochytrium* 23B (i.e., a "super-DHA-producer" *Schizochytrium*, wherein the production capabilities of the *Schizochytrium* combine with the DHA:DPA ratio of *Thraustochytrium*).

Therefore, the present invention makes use of genes from Thraustochytrid PUFA PKS systems, and further utilizes gene mixing to extend and/or alter the range of PUFA products to include EPA, DHA, DPA, ARA, GLA, SDA and others. The method to obtain

these altered PUFA production profiles includes not only the mixing of genes from various organisms into the Thrasustochytrid PUFA PKS genes, but also various methods of genetically modifying the endogenous Thraustochytrid PUFA PKS genes disclosed herein. Knowledge of the genetic basis and domain structure of the Thraustochytrid PUFA PKS system of the present invention (e.g., described in detail for *Schizochytrium* above) provides a basis for designing novel genetically modified organisms which produce a variety of PUFA profiles. Novel PUFA PKS constructs prepared in microorganisms such as a Thraustochytrid can be isolated and used to transform plants to impart similar PUFA production properties onto the plants.

Any one or more of the endogenous Thraustochytrid PUFA PKS domains can be altered or replaced according to the present invention, provided that the modification produces the desired result (i.e., alteration of the PUFA production profile of the microorganism). Particularly preferred domains to alter or replace include, but are not limited to, any of the domains corresponding to the domains in *Schizochytrium* OrfB or OrfC (β-keto acyl-ACP synthase (KS), acyltransferase (AT), FabA-like β-hydroxy acyl-ACP dehydrase (DH), chain length factor (CLF), enoyl ACP-reductase (ER), an enzyme that catalyzes the synthesis of *trans*-2-acyl-ACP, an enzyme that catalyzes the reversible isomerization of *trans*-2-acyl-ACP to *cis*-3-acyl-ACP, and an enzyme that catalyzes the elongation of *cis*-3-acyl-ACP to *cis*-5-β-keto-acyl-ACP). In one embodiment, preferred domains to alter or replace include, but are not limited to, β-keto acyl-ACP synthase (KS), FabA-like β-hydroxy acyl-ACP dehydrase (DH), and chain length factor (CLF).

In one aspect of the invention, Thraustochytrid PUFA-PKS PUFA production is altered by modifying the CLF (chain length factor) domain. This domain is characteristic of Type II (dissociated enzymes) PKS systems. Its amino acid sequence shows homology to KS (keto synthase pairs) domains, but it lacks the active site cysteine. CLF may function to determine the number of elongation cycles, and hence the chain length, of the end product. In this embodiment of the invention, using the current state of knowledge of FAS and PKS synthesis, a rational strategy for production of ARA by directed modification of the non-bacterial PUFA-PKS system is provided. There is controversy in the literature concerning

the function of the CLF in PKS systems (Bisang et al., *Nature* **401**, 502 (1999); Yi et al., *J. Am. Chem. Soc.* **125**, 12708 (2003)) and it is realized that other domains may be involved in determination of the chain length of the end product. However, it is significant that *Schizochytrium* produces both DHA (C22:6, ω-3) and DPA (C22:5, ω-6). In the PUFA-PKS system the *cis* double bonds are introduced during synthesis of the growing carbon chain. Since placement of the ω-3 and ω-6 double bonds occurs early in the synthesis of the molecules, one would not expect that they would affect subsequent end-product chain length determination. Thus, without being bound by theory, the present inventors believe that introduction of a factor (e.g. CLF) that directs synthesis of C20 units (instead of C22 units) into the *Schizochytrium* PUFA-PKS system will result in the production of EPA (C20:5, ω-3) and ARA (C20:4, ω-6). For example, in heterologous systems, one could exploit the CLF by directly substituting a CLF from an EPA producing system (such as one from *Photobacterium*, or preferably from a microorganism with the preferred growth requirements as described below) into the *Schizochytrium* gene set. The fatty acids of the resulting transformants can then be analyzed for alterations in profiles to identify the transformants producing EPA and/or ARA.

By way of example, in this aspect of the invention, one could construct a clone with the CLF of OrfB replaced with a CLF from a C20 PUFA-PKS system. A marker gene could be inserted downstream of the coding region. More specifically, one can use the homologous recombination system for transformation of Thraustochytrids as described herein and in detail in U.S. Patent Application Serial No. 10/124,807, *supra*. One can then transform the wild type Thraustochytrid cells (e.g., *Schizochytrium* cells), select for the marker phenotype, and then screen for those that had incorporated the new CLF. Again, one would analyze these transformants for any effects on fatty acid profiles to identify transformants producing EPA and/or ARA. If some factor other than those associated with the CLF is found to influence the chain length of the end product, a similar strategy could be employed to alter those factors.

In another aspect of the invention, modification or substitution of the β-hydroxy acyl-ACP dehydrase/keto synthase pairs is contemplated. During *cis*-vaccenic acid (C18:1, Δ11)

synthesis in *E. coli*, creation of the *cis* double bond is believed to depend on a specific DH enzyme, β-hydroxy acyl-ACP dehydrase, the product of the *fabA* gene. This enzyme removes HOH from a β-keto acyl-ACP and leaves a *trans* double bond in the carbon chain. A subset of DH's, FabA-like, possess *cis-trans* isomerase activity (Heath et al., 1996, *supra*). A novel aspect of bacterial and non-bacterial PUFA-PKS systems is the presence of two FabA-like DH domains. Without being bound by theory, the present inventors believe that one or both of these DH domains will possess *cis-trans* isomerase activity (manipulation of the DH domains is discussed in greater detail below).

Another aspect of the unsaturated fatty acid synthesis in *E. coli* is the requirement for a particular KS enzyme, β-ketoacyl-ACP synthase, the product of the *fabB* gene. This is the enzyme that carries out condensation of a fatty acid, linked to a cysteine residue at the active site (by a thio-ester bond), with a malonyl-ACP. In the multi-step reaction, CO_2 is released and the linear chain is extended by two carbons. It is believed that only this KS can extend a carbon chain that contains a double bond. This extension occurs only when the double bond is in the *cis* configuration; if it is in the *trans* configuration, the double bond is reduced by enoyl-ACP reductase (ER) prior to elongation (Heath et al., 1996, *supra*). All of the PUFA-PKS systems characterized so far have two KS domains, one of which shows greater homology to the FabB-like KS of *E. coli* than the other. Again, without being bound by theory, the present inventors believe that in PUFA-PKS systems, the specificities and interactions of the DH (FabA-like) and KS (FabB-like) enzymatic domains determine the number and placement of *cis* double bonds in the end products. Because the number of 2-carbon elongation reactions is greater than the number of double bonds present in the PUFA-PKS end products, it can be determined that in some extension cycles complete reduction occurs. Thus the DH and KS domains can be used as targets for alteration of the DHA/DPA ratio or ratios of other long chain fatty acids. These can be modified and/or evaluated by introduction of homologous domains from other systems or by mutagenesis of these gene fragments.

In another embodiment, the ER (enoyl-ACP reductase - an enzyme which reduces the trans-double bond in the fatty acyl-ACP resulting in fully saturated carbons) domains can be

modified or substituted to change the type of product made by the PKS system. For example, the present inventors know that *Schizochytrium* PUFA-PKS system differs from the previously described bacterial systems in that it has two (rather than one) ER domains. Without being bound by theory, the present inventors believe these ER domains can strongly influence the resulting PKS production product. The resulting PKS product could be changed by separately knocking out the individual domains or by modifying their nucleotide sequence or by substitution of ER domains from other organisms.

In another aspect of the invention, substitution of one of the DH (FabA-like) domains of the PUFA-PKS system for a DH domain that does not posses isomerization activity is contemplated, potentially creating a molecule with a mix of *cis*- and *trans*- double bonds. The current products of the *Schizochytrium* PUFA PKS system are DHA and DPA (C22:5 ω6). If one manipulated the system to produce C20 fatty acids, one would expect the products to be EPA and ARA (C20:4 ω6). This could provide a new source for ARA. One could also substitute domains from related PUFA-PKS systems that produced a different DHA to DPA ratio – for example by using genes from *Thraustochytrium* 23B (the PUFA PKS system of which is identified in U.S. Patent Application Serial No. 10/124,800, *supra*).

Additionally, in one embodiment, one of the ER domains is altered in the Thraustochytrid PUFA PKS system (e.g. by removing or inactivating) to alter the end product profile. Similar strategies could be attempted in a directed manner for each of the distinct domains of the PUFA-PKS proteins using more or less sophisticated approaches. Of course one would not be limited to the manipulation of single domains. Finally, one could extend the approach by mixing domains from the PUFA-PKS system and other PKS or FAS systems (e.g., type I, type II, modular) to create an entire range of new PUFA end products.

It is recognized that many genetic alterations, either random or directed, which one may introduce into a native (endogenous, natural) PKS system, will result in an inactivation of enzymatic functions. Therefore, in order to test for the effects of genetic manipulation of a Thraustochytrid PUFA PKS system in a controlled environment, one could first use a recombinant system in another host, such as *E. coli*, to manipulate various aspects of the

system and evaluate the results. For example, the FabB- strain of *E. coli* is incapable of synthesizing unsaturated fatty acids and requires supplementation of the medium with fatty acids that can substitute for its normal unsaturated fatty acids in order to grow (see Metz et al., 2001, *supra*). However, this requirement (for supplementation of the medium) can be removed when the strain is transformed with a functional PUFA-PKS system (i.e. one that produces a PUFA product in the *E. coli* host - see (Metz et al., 2001, *supra*, Figure 2A). The transformed FabB- strain now requires a functional PUFA-PKS system (to produce the unsaturated fatty acids) for growth without supplementation. The key element in this example is that production of a wide range of unsaturated fatty acid will suffice (even unsaturated fatty acid substitutes such as branched chain fatty acids). Therefore, in another preferred embodiment of the invention, one could create a large number of mutations in one or more of the PUFA PKS genes disclosed herein, and then transform the appropriately modified FabB- strain (e.g. create mutations in an expression construct containing an ER domain and transform a FabB- strain having the other essential domains on a separate plasmid - or integrated into the chromosome) and select only for those transformants that grow without supplementation of the medium (i.e., that still possessed an ability to produce a molecule that could complement the FabB- defect).

One test system for genetic modification of a PUFA PKS is exemplified in the Examples section. Briefly, a host microorganism such as *E. coli* is transformed with genes encoding a PUFA PKS system including all or a portion of a Thraustochytrid PUFA PKS system (e.g., Orfs A, B and C of *Schizochytrium*) and a gene encoding a phosphopantetheinyl transferases (PPTase), which is required for the attachment of a phosphopantetheine cofactor to produce the active, holo-ACP in the PKS system. The genes encoding the PKS system can be genetically engineered to introduce one or more modifications to the Thraustochytrid PUFA PKS genes and/or to introduce nucleic acids encoding domains from other PKS systems into the Thraustochytrid genes (including genes from non-Thraustochytrid microorganisms and genes from different Thraustochytrid microorganisms). The PUFA PKS system can be expressed in the *E. coli* and the PUFA production profile measured. In this

manner, potential genetic modifications can be evaluated prior to manipulation of the Thraustochytrid PUFA production organism.

The present invention includes the manipulation of endogenous nucleic acid molecules and/or the use of isolated nucleic acid molecules comprising a nucleic acid sequence from a Thraustochytrid PUFA PKS system or a homologue thereof. In one aspect, the present invention relates to the modification and/or use of a nucleic acid molecule comprising a nucleic acid sequence encoding a domain from a PUFA PKS system having a biological activity of at least one of the following proteins: malonyl-CoA:ACP acyltransferase (MAT), β-keto acyl-ACP synthase (KS), ketoreductase (KR), acyltransferase (AT), FabA-like β-hydroxy acyl-ACP dehydrase (DH), phosphopantetheine transferase, chain length factor (CLF), acyl carrier protein (ACP), enoyl ACP-reductase (ER), an enzyme that catalyzes the synthesis of *trans*-2-acyl-ACP, an enzyme that catalyzes the reversible isomerization of *trans*-2-acyl-ACP to *cis*-3-acyl-ACP, and/or an enzyme that catalyzes the elongation of *cis*-3-acyl-ACP to *cis*-5-β-keto-acyl-ACP. Preferred domains to modify in order to alter the PUFA production profile of a host Thraustochytrid have been discussed previously herein.

The genetic modification of a Thraustochytrid microorganism according to the present invention preferably affects the type, amounts, and/or activity of the PUFAs produced by the microorganism, whether the endogenous PUFA PKS system is genetically modified and/or whether recombinant nucleic acid molecules are introduced into the organism. According to the present invention, to affect an activity of a PUFA PKS system, such as to affect the PUFA production profile, includes any genetic modification in the PUFA PKS system or genes that interact with the PUFA PKS system that causes any detectable or measurable change or modification in any biological activity the PUFA PKS system expressed by the organism as compared to in the absence of the genetic modification. According to the present invention, the phrases "PUFA profile", "PUFA expression profile" and "PUFA production profile" can be used interchangeably and describe the overall profile of PUFAs expressed/produced by a microorganism. The PUFA expression profile can include the types of PUFAs expressed by the microorganism, as well as the absolute and

relative amounts of the PUFAs produced. Therefore, a PUFA profile can be described in terms of the ratios of PUFAs to one another as produced by the microorganism, in terms of the types of PUFAs produced by the microorganism, and/or in terms of the types and absolute or relative amounts of PUFAs produced by the microorganism.

As discussed above, while the host microorganism can include any eukaryotic microorganism with an endogenous PUFA PKS system and the ability to efficiently channel the products of the PUFA PKS system into both the phospholipids (PL) and triacylglycerols (TAG), the preferred host microorganism is any member of the order Thraustochytriales, including the families Thraustochytriaceae and Labyrinthulaceae. Particularly preferred host cells for use in the present invention could include microorganisms from a genus including, but not limited to: *Thraustochytrium, Japonochytrium, Aplanochytrium, Elina*, and *Schizochytrium* within the Thraustochytriaceae, and *Labyrinthula, Labyrinthuloides*, and *Labyrinthomyxa* within the Labyrinthulaceae. Preferred species within these genera include, but are not limited to: any species within *Labyrinthula*, including *Labrinthula* sp., *Labyrinthula algeriensis, Labyrinthula cienkowskii, Labyrinthula chattonii, Labyrinthula coenocystis, Labyrinthula macrocystis, Labyrinthula macrocystis atlantica, Labyrinthula macrocystis macrocystis, Labyrinthula magnifica, Labyrinthula minuta, Labyrinthula roscoffensis, Labyrinthula valkanovii, Labyrinthula vitellina, Labyrinthula vitellina pacifica, Labyrinthula vitellina vitellina, Labyrinthula zopfii*; any *Labyrinthuloides* species, including *Labyrinthuloides* sp., *Labyrinthuloides minuta, Labyrinthuloides schizochytrops*; any *Labyrinthomyxa* species, including *Labyrinthomyxa* sp., *Labyrinthomyxa pohlia, Labyrinthomyxa sauvageaui*, any *Aplanochytrium* species, including *Aplanochytrium* sp. and *Aplanochytrium kerguelensis*; any *Elina* species, including *Elina* sp., *Elina marisalba, Elina sinorifica*; any *Japanochytrium* species, including *Japanochytrium* sp., *Japanochytrium marinum*; any *Schizochytrium* species, including *Schizochytrium* sp., *Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum, Schizochytrium octosporum;* and any *Thraustochytrium* species, including *Thraustochytrium* sp., *Thraustochytrium aggregatum, Thraustochytrium arudimentale, Thraustochytrium aureum, Thraustochytrium benthicola, Thraustochytrium globosum, Thraustochytrium kinnei,*

Thraustochytrium motivum, Thraustochytrium pachydermum, Thraustochytrium proliferum, Thraustochytrium roseum, Thraustochytrium striatum, Ulkenia sp., *Ulkenia minuta, Ulkenia profunda, Ulkenia radiate, Ulkenia sarkariana,* and *Ulkenia visurgensis.* Particularly preferred species within these genera include, but are not limited to: any *Schizochytrium* species, including *Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum*; any *Thraustochytrium* species (including former *Ulkenia* species such as *U. visurgensis, U. amoeboida, U. sarkariana, U. profunda, U. radiata, U. minuta* and *Ulkenia sp.* BP-5601), and including *Thraustochytrium striatum, Thraustochytrium aureum, Thraustochytrium roseum*; and any *Japonochytrium* species. Particularly preferred strains of Thraustochytriales include, but are not limited to: *Schizochytrium sp.* (S31)(ATCC 20888); *Schizochytrium sp.* (S8)(ATCC 20889); *Schizochytrium sp.* (LC-RM)(ATCC 18915); *Schizochytrium sp.* (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky)(ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi)(IFO 32693); *Thraustochytrium sp.* (23B)(ATCC 20891); *Thraustochytrium striatum* (Schneider)(ATCC 24473); *Thraustochytrium aureum* (Goldstein)(ATCC 34304); *Thraustochytrium roseum* (Goldstein)(ATCC 28210); and *Japonochytrium sp.* (L1)(ATCC 28207).

In one embodiment of the present invention, it is contemplated that a mutagenesis program could be combined with a selective screening process to obtain a Thraustochytrid microorganism with the PUFA production profile of interest. The mutagenesis methods could include, but are not limited to: chemical mutagenesis, gene shuffling, switching regions of the genes encoding specific enzymatic domains, or mutagenesis restricted to specific regions of those genes, as well as other methods.

For example, high throughput mutagenesis methods could be used to influence or optimize production of the desired PUFA profile. Once an effective model system has been developed, one could modify these genes in a high throughput manner. Utilization of these technologies can be envisioned on two levels. First, if a sufficiently selective screen for production of a product of interest (e.g., EPA) can be devised, it could be used to attempt to alter the system to produce this product (e.g., in lieu of, or in concert with, other strategies such as those discussed above). Additionally, if the strategies outlined above resulted in a

set of genes that did produce the PUFA profile of interest, the high throughput technologies could then be used to optimize the system. For example, if the introduced domain only functioned at relatively low temperatures, selection methods could be devised to permit removing that limitation.

In one embodiment of the present invention, a genetically modified Thraustochytrid microorganism has an enhanced ability to synthesize desired PUFAs and/or has a newly introduced ability to synthesize a different profile of PUFAs. According to the present invention, "an enhanced ability to synthesize" a product refers to any enhancement, or up-regulation, in a pathway related to the synthesis of the product such that the microorganism produces an increased amount of the product (including any production of a product where there was none before) as compared to the wild-type microorganism, cultured or grown, under the same conditions. Methods to produce such genetically modified organisms have been described in detail above.

As described above, in one embodiment of the present invention, a genetically modified microorganism or plant includes a microorganism or plant which has an enhanced ability to synthesize desired bioactive molecules (products) or which has a newly introduced ability to synthesize specific products (e.g., to synthesize a specific antibiotic). According to the present invention, "an enhanced ability to synthesize" a product refers to any enhancement, or up-regulation, in a pathway related to the synthesis of the product such that the microorganism or plant produces an increased amount of the product (including any production of a product where there was none before) as compared to the wild-type microorganism or plant, cultured or grown, under the same conditions. Methods to produce such genetically modified organisms have been described in detail above.

One embodiment of the present invention is a method to produce desired bioactive molecules (also referred to as products or compounds) by growing or culturing a genetically modified microorganism or plant of the present invention (described in detail above). Such a method includes the step of culturing in a fermentation medium or growing in a suitable environment, such as soil, a microorganism or plant, respectively, that has a genetic modification as described previously herein and in accordance with the present invention.

Preferred host cells for genetic modification related to the PUFA PKS system of the invention are described above.

One embodiment of the present invention is a method to produce desired PUFAs by culturing a genetically modified Thraustochytrid microorganism of the present invention (described in detail above). Such a method includes the step of culturing in a fermentation medium and under conditions effective to produce the PUFA(s) a Thraustochytrid microorganism that has a genetic modification as described previously herein and in accordance with the present invention. An appropriate, or effective, medium refers to any medium in which a genetically modified microorganism of the present invention, including Thraustochytrids and other microorganisms, when cultured, is capable of producing the desired PUFA product(s). Such a medium is typically an aqueous medium comprising assimilable carbon, nitrogen and phosphate sources. Such a medium can also include appropriate salts, minerals, metals and other nutrients. Any microorganisms of the present invention can be cultured in conventional fermentation bioreactors. The microorganisms can be cultured by any fermentation process which includes, but is not limited to, batch, fed-batch, cell recycle, and continuous fermentation. Preferred growth conditions for Thraustochytrid microorganisms according to the present invention are well known in the art and are described in detail, for example, in U.S. Patent No. 5,130,242, U.S. Patent No. 5,340,742, and U.S. Patent No. 5,698,244, each of which is incorporated herein by reference in its entirety.

In one embodiment, the genetically modified microorganism is cultured at a temperature of greater than about 15°C, and in another embodiment, greater than about 20°C, and in another embodiment, greater than about 25°C, and in another embodiment, greater than about 30°C, and in another embodiment, greater than about 35°C, and in another embodiment, greater than about 40°C, and in one embodiment, at any temperature between about 20°C and 40°C.

The desired PUFA(s) and/or other bioactive molecules produced by the genetically modified microorganism can be recovered from the fermentation medium using conventional separation and purification techniques. For example, the fermentation medium can be

filtered or centrifuged to remove microorganisms, cell debris and other particulate matter, and the product can be recovered from the cell-free supernatant by conventional methods, such as, for example, ion exchange, chromatography, extraction, solvent extraction, phase separation, membrane separation, electrodialysis, reverse osmosis, distillation, chemical derivatization and crystallization. Alternatively, microorganisms producing the PUFA(s), or extracts and various fractions thereof, can be used without removal of the microorganism components from the product.

Preferably, a genetically modified Thraustochytrid microorganism of the invention produces one or more polyunsaturated fatty acids including, but not limited to, EPA (C20:5, ω-3), DHA (C22:6, ω-3), DPA (C22:5, ω-6), ARA (C20:4, ω-6), GLA (C18:3, n-6), and SDA (C18:4, n-3)). In one preferred embodiment, a *Schizochytrium* that, in wild-type form, produces high levels of DHA and DPA, is genetically modified according to the invention to produce high levels of EPA. As discussed above, one advantage of using genetically modified Thraustochytrid microorganisms to produce PUFAs is that the PUFAs are directly incorporated into both the phospholipids (PL) and triacylglycerides (TAG).

Preferably, PUFAs are produced in an amount that is greater than about 5% of the dry weight of the microorganism, and in one aspect, in an amount that is greater than 6%, and in another aspect, in an amount that is greater than 7%, and in another aspect, in an amount that is greater than 8%, and in another aspect, in an amount that is greater than 9%, and in another aspect, in an amount that is greater than 10%, and so on in whole integer percentages, up to greater than 90% dry weight of the microorganism (e.g., 15%, 20%, 30%, 40%, 50%, and any percentage in between).

In the method for production of desired bioactive compounds of the present invention, a genetically modified plant is cultured in a fermentation medium or grown in a suitable medium such as soil. An appropriate, or effective, fermentation medium has been discussed in detail above. A suitable growth medium for higher plants includes any growth medium for plants, including, but not limited to, soil, sand, any other particulate media that support root growth (e.g. vermiculite, perlite, etc.) or hydroponic culture, as well as suitable light, water and nutritional supplements which optimize the growth of the higher plant. The

genetically modified plants of the present invention are engineered to produce significant quantities of the desired product through the activity of the PKS system that is genetically modified according to the present invention. The compounds can be recovered through purification processes which extract the compounds from the plant. In a preferred embodiment, the compound is recovered by harvesting the plant. In this embodiment, the plant can be consumed in its natural state or further processed into consumable products.

Many genetic modifications useful for producing bioactive molecules will be apparent to those of skill in the art, given the present disclosure, and various other modifications have been discussed previously herein. The present invention contemplates any genetic modification related to a PUFA PKS system as described herein which results in the production of a desired bioactive molecule.

Bioactive molecules, according to the present invention, include any molecules (compounds, products, etc.) that have a biological activity, and that can be produced by a PKS system that comprises at least one amino acid sequence having a biological activity of at least one functional domain of a non-bacterial PUFA PKS system as described herein. Such bioactive molecules can include, but are not limited to: a polyunsaturated fatty acid (PUFA), an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and a cholesterol lowering formulation. One advantage of the non-bacterial PUFA PKS system of the present invention is the ability of such a system to introduce carbon-carbon double bonds in the *cis* configuration, and molecules including a double bond at every third carbon. This ability can be utilized to produce a variety of compounds.

Preferably, bioactive compounds of interest are produced by the genetically modified microorganism in an amount that is greater than about 0.05%, and preferably greater than about 0.1%, and more preferably greater than about 0.25%, and more preferably greater than about 0.5%, and more preferably greater than about 0.75%, and more preferably greater than about 1%, and more preferably greater than about 2.5%, and more preferably greater than

about 5%, and more preferably greater than about 10%, and more preferably greater than about 15%, and even more preferably greater than about 20% of the dry weight of the microorganism. For lipid compounds, preferably, such compounds are produced in an amount that is greater than about 5% of the dry weight of the microorganism. For other bioactive compounds, such as antibiotics or compounds that are synthesized in smaller amounts, those strains possessing such compounds at of the dry weight of the microorganism are identified as predictably containing a novel PKS system of the type described above. In some embodiments, particular bioactive molecules (compounds) are secreted by the microorganism, rather than accumulating. Therefore, such bioactive molecules are generally recovered from the culture medium and the concentration of molecule produced will vary depending on the microorganism and the size of the culture.

One embodiment of the present invention relates to a method to modify an endproduct containing at least one fatty acid, comprising adding to the endproduct an oil produced by a recombinant host cell that expresses at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system. The PUFA PKS system includes any suitable bacterial or non-bacterial PUFA PKS system described herein, including the PUFA PKS systems from *Thraustochytrium* and *Schizochytrium*, or any PUFA PKS system from bacteria that normally (i.e., under normal or natural conditions) are capable of growing and producing PUFAs at temperatures above 22°C, such as *Shewanella olleyana* or *Shewanella japonica*.

Preferably, the endproduct is selected from the group consisting of a food, a dietary supplement, a pharmaceutical formulation, a humanized animal milk, and an infant formula. Suitable pharmaceutical formulations include, but are not limited to, an anti-inflammatory formulation, a chemotherapeutic agent, an active excipient, an osteoporosis drug, an anti-depressant, an anti-convulsant, an anti-*Heliobactor pylori* drug, a drug for treatment of neurodegenerative disease, a drug for treatment of degenerative liver disease, an antibiotic, and a cholesterol lowering formulation. In one embodiment, the endproduct is used to treat a condition selected from the group consisting of: chronic inflammation, acute inflammation, gastrointestinal disorder, cancer, cachexia, cardiac restenosis, neurodegenerative disorder,

degenerative disorder of the liver, blood lipid disorder, osteoporosis, osteoarthritis, autoimmune disease, preeclampsia, preterm birth, age related maculopathy, pulmonary disorder, and peroxisomal disorder.

Suitable food products include, but are not limited to, fine bakery wares, bread and rolls, breakfast cereals, processed and unprocessed cheese, condiments (ketchup, mayonnaise, etc.), dairy products (milk, yogurt), puddings and gelatin desserts, carbonated drinks, teas, powdered beverage mixes, processed fish products, fruit-based drinks, chewing gum, hard confectionery, frozen dairy products, processed meat products, nut and nut-based spreads, pasta, processed poultry products, gravies and sauces, potato chips and other chips or crisps, chocolate and other confectionery, soups and soup mixes, soya based products (milks, drinks, creams, whiteners), vegetable oil-based spreads, and vegetable-based drinks.

Yet another embodiment of the present invention relates to a method to produce a humanized animal milk. This method includes the steps of genetically modifying milk-producing cells of a milk-producing animal with at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding at least one biologically active domain of a PUFA PKS system as described herein.

Methods to genetically modify a host cell and to produce a genetically modified non-human, milk-producing animal, are known in the art. Examples of host animals to modify include cattle, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of a transgene expressing population. For animals, PKS-like transgenes can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUFAs in the breast milk of the host animal.

The following examples are provided for the purpose of illustration and are not intended to limit the scope of the present invention.

Examples

Example 1

The following example, from U.S. Patent Application No. 10/124,800, describes the use of the screening process of the present invention to identify other non-bacterial organisms comprising a PUFA PKS system according to the present invention.

Thraustochytrium sp. 23B (ATCC 20892) was cultured as described in detail herein.

A frozen vial of *Thraustochytrium* sp. 23B (ATCC 20892) was used to inoculate a 250 mL shake flask containing 50 mL of RCA medium. The culture was shaken on a shaker table (200 rpm) for 72 hr at 25°C. RCA medium contains the following:

RCA Medium

Deionized water	1000 mL
Reef Crystals® sea salts	40 g/L
Glucose	20 g/L
Monosodium glutamate (MSG)	20 g/L
Yeast extract	1 g/L
PII metals*	5 mL/L
Vitamin mix*	1 mL/L
pH	7.0

*PII metal mix and vitamin mix are same as those outlined in U.S. Patent No. 5,130,742, incorporated herein by reference in its entirety.

25 mL of the 72 hr old culture was then used to inoculate another 250 mL shake flask containing 50 mL of low nitrogen RCA medium (10 g/L MSG instead of 20 g/L) and the other 25 mL of culture was used to inoculate a 250 mL shake flask containing 175 mL of low-nitrogen RCA medium. The two flasks were then placed on a shaker table (200 rpm) for 72 hr at 25°C. The cells were then harvested via centrifugation and dried by lyophilization. The dried cells were analyzed for fat content and fatty acid profile and content using standard gas chromatograph procedures.

The screening results for *Thraustochytrium* 23B under low oxygen conditions relative to high oxygen conditions were as follows:

Did DHA as % FAME increase?	Yes (38->44%)
C14:0 + C16:0 + C16:1 greater than about 40% TFA?	Yes (44%)

No C18:3(n-3) or C18:3(n-6)?	Yes (0%)
Did fat content increase?	Yes (2-fold increase)
Did DHA (or other HUFA content increase)?	Yes (2.3-fold increase)

The results, especially the significant increase in DHA content (as % FAME) under low oxygen conditions, conditions, strongly indicates the presence of a PUFA producing PKS system in this strain of *Thraustochytrium*.

In order to provide additional data confirming the presence of a PUFA PKS system, a Southern blot of *Thraustochytrium* 23B was conducted using PKS probes from *Schizochytrium* strain 20888, a strain which has already been determined to contain a PUFA producing PKS system (i.e., SEQ ID Nos:1-32 described above). Fragments of *Thraustochytrium* 23B genomic DNA which are homologous to hybridization probes from PKS PUFA synthesis genes were detected using the Southern blot technique. *Thraustochytrium* 23B genomic DNA was digested with either *Cla*I or *Kpn*I restriction endonucleases, separated by agarose gel electrophoresis (0.7% agarose, in standard tris-acetate-EDTA buffer), and blotted to a Schleicher &Schuell Nytran Supercharge membrane by capillary transfer. Two digoxigenin labeled hybridization probes were used - one specific for the enoyl-ACP reductase (ER) region of *Schizochytrium* PKS Orf B (nucleotides 5012-5511 of Orf B; SEQ ID NO:3), and the other specific for a conserved region at the beginning of *Schizochytrium* PKS Orf C (nucleotides 76-549 of OrfC; SEQ ID NO:5).

The OrfB-ER probe detected an approximately 13kb ClaI fragment and an approximately 3.6 kb KpnI fragment in the *Thraustochytrium* 23B genomic DNA. The OrfC probe detected an approximately 7.5 kb ClaI fragment and an approximately 4.6 kb KpnI fragment in the *Thraustochytrium* 23B genomic DNA.

Finally, a recombinant genomic library, consisting of DNA fragments from *Thraustochytrium* 23B genomic DNA inserted into vector lambda FIX II (Stratagene), was screened using digoxigenin labeled probes corresponding to the following segments of *Schizochytrium* 20888 PUFA-PKS genes: nucleotides 7385-7879 of Orf A (SEQ ID NO:1), nucleotides 5012-5511 of Orf B (SEQ ID NO:3), and nucleotides 76-549 of Orf C (SEQ ID NO:5). Each of these probes detected positive plaques from the *Thraustochytrium* 23B

library, indicating extensive homology between the *Schizochytrium* PUFA-PKS genes and the genes of *Thraustochytrium* 23B.

These results demonstrate that *Thraustochytrium* 23B genomic DNA contains sequences that are homologous to PKS genes from *Schizochytrium* 20888.

Example 2

The following example demonstrates that *Schizochytrium* Orfs A, B and C encode a functional DHA/DPA synthesis enzyme via functional expression in *E. coli*.

General preparation of E. coli transformants

The three genes encoding the *Schizochytrium* PUFA PKS system that produces DHA and DPA in *Schizochytrium* (Orfs A, B & C; SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, respectively) were cloned into a single *E. coli* expression vector (derived from pET21c (Novagen)). The genes are transcribed as a single message (by the T7 RNA-polymerase), and a ribosome-binding site cloned in front of each of the genes initiates translation. Modification of the Orf B coding sequence was needed to obtain production of a full-length Orf B protein in *E. coli* (see below). An accessory gene, encoding a PPTase (see below) was cloned into a second plasmid (derived from pACYC184, New England Biolabs).

OrfB

The Orf B gene is predicted to encode a protein with a mass of ~224 kDa. Initial attempts at expression of the gene in *E. coli* resulted in accumulation of a protein with an apparent molecular mass of ~165 kDa (as judged by comparison to proteins of known mass during SDS-PAGE). Examination of the Orf B nucleotide sequence revealed a region containing 15 sequential serine codons – all of them being the TCT codon. The genetic code contains 6 different serine codons, and three of these are used frequently in *E. coli*. The present inventors used four overlapping oligonucleotides in combination with a polymerase chain reaction protocol to resynthesize a small portion of the Orf B gene (a ~195 base pair, BspHI to SacII restriction enzyme fragment) that contained the serine codon repeat region. In the synthetic Orf B fragment, a random mixture of the 3 serine codons commonly used by *E. coli* was used, and some other potentially problematic codons were changed as well (i.e., other codons rarely used by *E. coli*). The BspHI to SacII fragment present in the original Orf

B was replaced by the resynthesized fragment (to yield Orf B*) and the modified gene was cloned into the relevant expression vectors. The modified OrfB* still encodes the amino acid sequence of SEQ ID NO:4. Expression of the modified Orf B* clone in *E. coli* resulted in the appearance of a ~224 kDa protein, indicating that the full-length product of OrfB was produced. The sequence of the resynthesized Orf B* BspHI to SacII fragment is shown in SEQ ID NO:80. Referring to SEQ ID NO:80, the nucleotide sequence of the resynthesized BspHI to SacII region of Orf B is shown. The BspHI restriction site and the SacII restriction site are identified. The BspHI site starts at nucleotide 4415 of the Orf B CDS (SEQ ID NO:3) (note: there are a total of three BspHI sites in the Orf B CDS, while the SacII site is unique). The sequence of the unmodified Orf B CDS is given in GenBank Accession number AF378328 and in SEQ ID NO:3.

PPTase

The ACP domains of the Orf A protein (SEQ ID NO:2 in *Schizochytrium*) must be activated by addition of phosphopantetheine group in order to function. The enzymes that catalyze this general type of reaction are called phosphopantetheine transferases (PPTases). *E. coli* contains two endogenous PPTases, but it was anticipated that they would not recognize the Orf A ACP domains from *Schizochytrium*. This was confirmed by expressing Orfs A, B* (see above) and C in *E. coli* without an additional PPTase. In this transformant, no DHA production was detected. The inventors tested two heterologous PPTases in the *E. coli* PUFA PKS expression system: (1) sfp (derived from *Bacillus subtilis*) and (2) Het I (from the cyanobacterium *Nostoc* strain 7120).

The sfp PPTase has been well characterized and is widely used due to its ability to recognize a broad range of substrates. Based on published sequence information (Nakana, et al., 1992, *Molecular and General Genetics* 232: 313-321), an expression vector for sfp was built by cloning the coding region, along with defined up- and downstream flanking DNA sequences, into a pACYC-184 cloning vector. The oligonucleotides:

CGGGGTACCCGGGAGCCGCCTTGGCTTTGT (forward; SEQ ID NO:73); and

AAACTGCAGCCCGGGTCCAGCTGGCAGGCACCCTG (reverse; SEQ ID NO:74),

were used to amplify the region of interest from genomic *B. subtilus* DNA. Convenient restriction enzyme sites were included in the oligonucleotides to facilitate cloning in an intermediate, high copy number vector and finally into the EcoRV site of pACYC184 to create the plasmid: pBR301. Examination of extracts of *E. coli* transformed with this plasmid revealed the presence of a novel protein with the mobility expected for sfp. Co-expression of the sfp construct in cells expressing the Orf A, B*, C proteins, under certain conditions, resulted in DHA production. This experiment demonstrated that sfp was able to activate the *Schizochytrium* Orf A ACP domains. In addition, the regulatory elements associated with the sfp gene were used to create an expression cassette into which other genes could be inserted. Specifically, the sfp coding region (along with three nucleotides immediately upstream of the ATG) in pBR301 was replaced with a 53 base pair section of DNA designed so that it contains several unique (for this construct) restriction enzyme sites. The initial restriction enzyme site in this region is NdeI (CATATG; SEQ ID NO:79). The ATG sequence embedded in this site is utilized as the initiation methionine codon for introduced genes. The additional restriction sites (BglLL, NotI, SmaI, PmeII, HindIII, SpeI and XhoI) were included to facilitate the cloning process. The functionality of this expression vector cassette was tested by using PCR to generate a version of sfp with a NdeI site at the 5' end and an XhoI site ate the 3' end. This fragment was cloned into the expression cassette and transferred into *E. coli* along with the Orf A, B* and C expression vector. Under appropriate conditions, these cells accumulated DHA, demonstrating that a functional sfp had been produced.

To the present inventors' knowledge, Het I has not been tested previously in a heterologous situation. Het I is present in a cluster of genes in Nostoc known to be responsible for the synthesis of long chain hydroxy-fatty acids that are a component of a glyco-lipid layer present in heterocysts of that organism. The present inventors, without being bound by theory, believe that Het I activates the ACP domains of a protein, Hgl E, present in that cluster. The two ACP domains of Hgl E have a high degree of sequence

homology to the ACP domains found in *Schizochytrium* Orf A. The endogenous start codon of Het I has not been identified (there is no methionine present in the putative protein). There are several potential alternative start codons (e.g., TTG and ATT) near the 5' end of the open reading frame. The sequence of the region of Nostoc DNA encoding the HetI gene is shown in SEQ ID NO:81. SEQ ID NO:82 represents the amino acid sequence encoded by SEQ ID NO:81. Referring to SEQ ID NO:81, limit to the upstream coding region indicated by the inframe nonsense triplet (TAA) at positions 1-3 of SEQ ID NO:81 and ends with the stop codon (TGA) at positions 715-717 of SEQ ID NO:81. No methionine codons (ATG) are present in the sequence. Potential alternative initiation codons are: 3 TTG codons (positions 4-6, 7-9 and 49-51 of SEQ ID NO:81), ATT (positions 76-78 of SEQ ID NO:81) and GTG (positions 235-237 of SEQ ID NO:81). A Het I expression construct was made by using PCR to replace the furthest 5' potential alternative start codon (TTG) with a methionine codon (ATG, as part of the above described NdeI restriction enzyme recognition site), and introducing an XhoI site at the 3' end of the coding sequence. The modified HetI coding sequence was then inserted into the NdeI and XhoI sites of the pACYC184 vector construct containing the sfp regulatory elements. Expression of this Het I construct in *E. coli* resulted in the appearance of a new protein of the size expected from the sequence data. Co-expression of Het I with *Schizochytrium* Orfs A, B*, C in *E. coli* under several conditions resulted in the accumulation of DHA and DPA in those cells. In all of the experiments in which sfp and Het I were compared, more DHA and DPA accumulated in the cells containing the Het I construct than in cells containing the sfp construct.

Production of DHA and DPA in E. coli transformants

The two plasmids encoding: (1) the *Schizochytrium* PUFA PKS genes (Orfs A, B* and C) and (2) the PPTase (from sfp or from Het I) were transformed into *E. coli* strain BL21 which contains an inducible T7 RNA polymerase gene. Synthesis of the *Schizochytrium* proteins was induced by addition of IPTG to the medium, while PPTase expression was controlled by a separate regulatory element (see above). Cells were grown under various defined conditions and using either of the two heterologous PPTase genes. The cells were

harvested and the fatty acids were converted to methyl-esters (FAME) and analyzed using gas-liquid chromatography.

Under several conditions, DHA and DPA were detected in *E. coli* cells expressing the *Schizochytrium* PUFA PKS genes, plus either of the two heterologous PPTases. No DHA or DPA was detected in FAMEs prepared from control cells (i.e., cells transformed with a plasmid lacking one of the Orfs). The ratio of DHA to DPA observed in *E. coli* approximates that of the endogenous DHA and DPA production observed in *Schizochytrium*. The highest level of PUFA (DHA plus DPA), representing ~17% of the total FAME, was found in cells grown at 32°C in 765 medium (recipe available from the American Type Culture Collection) supplemented with 10% (by weight) glycerol. Note that PUFA accumulation was also observed when cells were grown in Luria Broth supplemented with 5 or 10 % glycerol, and when grown at 20°C. Selection for the presence of the respective plasmids was maintained by inclusion of the appropriate antibiotics during the growth and IPTG (to a final concentration of 0.5 mM) was used to induce expression of Orfs A, B* and C.

Fig. 4 shows an example chromatogram from gas-liquid chromatographic analysis of FAMEs derived from control cells and from cells expressing the *Schizochytrium* PUFA PKS genes plus a PPTase (in this case Het I). Identity of the labeled FAMEs has been confirmed using mass spectroscopy.

Example 3

The following example shows demonstrates that genes encoding the *Schizochytrium* PUFA PKS enzyme complex can be selectively inactivated (knocked out), and that it is a lethal phenotype unless the medium is supplemented with polyunsaturated fatty acids.

Homologous recombination has been demonstrated in *Schizochytrium* (see copending U.S. Patent Application Serial No. 10/124,807, incorporated herein by reference in its entirety). A plasmid designed to inactivate *Schizochytrium* Orf A (SEQ ID NO:1) was made by inserting a Zeocin™ resistance marker into the Sma I site of a clone containing the Orf A coding sequence. The Zeocin™ resistance marker was obtained from the plasmid pMON50000 - expression of the Zeocin™ resistance gene is driven by a *Schizochytrium*

derived tubulin promoter element (see U.S. Patent Application Serial No. 10/124,807, *ibid.*). The knock-out construct thus consists of: 5' *Schizochytrium* Orf A coding sequence, the tub-Zeocin™ resistance element and 3' *Schizochytrium* Orf A coding sequence, all cloned into pBluescript II SK (+) vector (Stratagene).

The plasmid was introduced into *Schizochytrium* cells by particle bombardment and transformants were selected on plates containing Zeocin™ and supplemented with polyunsaturated fatty acids (PUFA) (see Example 4). Colonies that grew on the Zeocin™ plus PUFA plates were tested for ability to grow on plates without the PUFA supplementation and several were found that required the PUFA. These PUFA auxotrophs are putative Orf A knockouts. Northern blot analysis of RNA extracted from several of these mutants confirmed that a full-length Orf A message was not produced in these mutants.

These experiments demonstrate that a *Schizochytrium* gene (e.g., Orf A) can be inactivated via homologous recombination, that inactivation of Orf A results in a lethal phenotype, and that those mutants can be rescued by supplementation of the media with PUFA.

Similar sets of experiments directed to the inactivation of *Schizochytrium* Orf B (SEQ ID NO:3) and Orf C (SEQ ID NO:5) have yielded similar results. That is, Orf B and Orf C can be individually inactivated by homologous recombination and those cells require PUFA supplementation for growth.

Example 4

The following example shows that PUFA auxotrophs can be maintained on medium supplemented with EPA, demonstrating that EPA can substitute for DHA in *Schizochytrium*.

As indicated in Example 3, *Schizochytrium* cells in which the PUFA PKS complex has been inactivated required supplementation with PUFA to survive. Aside from demonstrating that *Schizochytrium* is dependent on the products of this system for growth, this experimental system permits the testing of various fatty acids for their ability to rescue the mutants. It was discovered that the mutant cells (in which any of the three genes have been inactivated) grew as well on media supplemented with EPA as they did on media supplemented with DHA. This result indicates that, if the endogenous PUFA PKS complex

which produces DHA were replaced with one whose product was EPA, the cells would be viable. Additionally, these mutant cells could be rescued by supplementation with either ARA or GLA, demonstrating the feasibility of producing genetically modified *Schizochytrium* that produce these products. It is noted that a preferred method for supplementation with PUFAs involves combining the free fatty acids with partially methylated beta-cyclodextrin prior to addition of the PUFAs to the medium.

Example 5

The following example shows that inactivated PUFA genes can be replaced at the same site with active forms of the genes in order to restore PUFA synthesis.

Double homologous recombination at the acetolactate synthase gene site has been demonstrated in *Schizochytrium* (see U.S. Patent Application Serial No. 10/124,807, *supra*). The present inventors tested this concept for replacement of the *Schizochytrium* PUFA PKS genes by transformation of a *Schizochytrium* Orf A knockout strain (described in Example 2) with a full-length *Schizochytrium* Orf A genomic clone. The transformants were selected by their ability to grow on media without supplemental PUFAs. These PUFA prototrophs were then tested for resistance to Zeocin™ and several were found that were sensitive to the antibiotic. These results indicate that the introduced *Schizochytrium* Orf A has replaced the Zeocin™ resistance gene in the knockout strain via double homologous recombination. This experiment demonstrates the proof of concept for gene replacement within the PUFA PKS genes. Similar experiments for *Schizochytrium* Orf B and Orf C knock-outs have given identical results.

Example 6

This example shows that all or some portions of the *Thraustochytrium* 23B PUFA PKS genes can function in *Schizochytrium*.

As described in U.S Patent Application Serial No. 10/124,800 (*supra*), the DHA-producing protist *Thraustochytrium* 23B (*Th.* 23B) has been shown to contain orfA, orfB, and orfC homologs. Complete genomic clones of the three *Th.* 23B genes were used to transform the *Schizochytrium* strain containing the cognate orf "knock-out". Direct selection for complemented transformants was carried out in the absence of PUFA supplementation.

By this method, it was shown that the *Th.* 23B orfA and orfC genes could complement the *Schizochytrium* orfA and orfC knock-out strains, respectively, to PUFA prototrophy. Complemented transformants were found that either retained or lost Zeocin™ resistance (the marker inserted into the *Schizochytrium* genes thereby defining the knock-outs). The Zeocin™-resistant complemented transformants are likely to have arisen by a single cross-over integration of the entire *Thraustochytrium* gene into the *Schizochytrium* genome outside of the respective orf region. This result suggests that the entire *Thraustochytrium* gene is functioning in *Schizochytrium*. The Zeocin™-sensitive complemented transformants are likely to have arisen by double cross-over events in which portions (or conceivably all) of the *Thraustochytrium* genes functionally replaced the cognate regions of the *Schizochytrium* genes that had contained the disruptive Zeocin™ resistance marker. This result suggests that a fraction of the *Thraustochytrium* gene is functioning in *Schizochytrium*.

Example 7

The following example shows that certain EPA-producing bacteria contain PUFA PKS-like genes that appear to be suitable for modification of *Schizochytrium*.

Two EPA-producing marine bacterial strains of the genus *Shewanella* have been shown to grow at temperatures typical of *Schizochytrium* fermentations and to possess PUFA PKS-like genes. *Shewanella olleyana* (Australian Collection of Antarctic Microorganisms (ACAM) strain number 644; Skerratt et al., *Int. J. Syst. Evol. Microbiol* **52**, 2101 (2002)) produces EPA and grows up to 30°C. *Shewanella japonica* (American Type Culture Collection (ATCC) strain number BAA-316; Ivanova et al., *Int. J. Syst. Evol. Microbiol.* **51**, 1027 (2001)) produces EPA and grows up to 35°C.

To identify and isolate the PUFA-PKS genes from these bacterial strains, degenerate PCR primer pairs for the KS-MAT region of bacterial orf5/pfaA genes and the DH-DH region of bacterial orf7/pfaC genes were designed based on published gene sequences for *Shewanella* SCRC-2738, *Shewanella oneidensis* MR-1; *Shewanella* sp. GA-22; *Photobacter profundum*, and *Moritella marina* (see discussion above). Specifically, the primers and PCR conditions were designed as follows:

Primers for the KS/AT region; based on the following published sequences: *Shewanella* sp. SCRC-2738; *Shewanella oneidensis* MR-1; *Photobacter profundum*; *Moritella marina*:

prRZ23 GGYATGMTGRTTGGTGAAGG (forward; SEQ ID NO:69)

prRZ24 TRTTSASRTAYTGYGAACCTTG (reverse; SEQ ID NO:70)

Primers for the DH region; based on the following published sequences: *Shewanella* sp. GA-22; *Shewanella* sp. SCRC-2738; *Photobacter profundum*; *Moritella marina*:

prRZ28 ATGKCNGAAGGTTGTGGCCA (forward; SEQ ID NO:71)

prRZ29 CCWGARATRAAGCCRTTDGGTTG (reverse; SEQ ID NO:72)

The PCR conditions (with bacterial chromosomal DNA as templates) were as follows:

Reaction Mixture:

0.2 μM dNTPs
0.1 μM each primer
8% DMSO
250 ng chromosomal DNA
2.5U Herculase® DNA polymerase (Stratagene)
1X Herculase® buffer
50μL total volume

PCR Protocol: (1) 98°C for 3 min.; (2) 98°C for 40 sec.; (3) 56°C for 30 sec.; (4) 72°C for 90 sec.; (5) Repeat steps 2-4 for 29 cycles; (6) 72°C for 10 min.; (7) Hold at 6°C.

For both primer pairs, PCR gave distinct products with expected sizes using chromosomal DNA templates from either *Shewanella olleyana* or *Shewanella japonica*. The four respective PCR products were cloned into pCR-BLUNT II-TOPO (Invitrogen) and insert sequences were determined using the M13 forward and reverse primers. In all cases, the DNA sequences thus obtained were highly homologous to known bacterial PUFA PKS gene regions.

The DNA sequences obtained from the bacterial PCR products were compared with known sequences and with PUFA PKS genes from *Schizochytrium* ATCC 20888 in a standard Blastx search (BLAST parameters: Low Complexity filter: On; Matrix: BLOSUM62; Word Size: 3; Gap Costs: Existance11, Extension 1 (BLAST described in Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporated herein by reference in its entirety)).

At the amino acid level, the sequences with the greatest degree of homology to the *Shewanella olleyana* ACAM644 ketoacyl synthase/acyl transferase (KS-AT) deduced amino acid sequence encoded by SEQ ID NO:76 were: *Photobacter profundum* pfaA (identity = 70%; positives = 81%); *Shewanella oneidensis* MR-1 "multi-domain β-ketoacyl synthase" (identity = 66%; positives = 77%); and *Moritella marina* ORF8 (identity = 56%; positives = 71%). The *Schizochytrium* sp. ATCC20888 orfA was 41% identical and 56% positive to the deduced amino acid sequence encoded by SEQ ID NO:76.

At the amino acid level, the sequences with the greatest degree of homology to the *Shewanella japonica* ATCC BAA-316 ketoacyl synthase/acyl transferase (KS-AT) deduced amino acid sequence encoded by SEQ ID NO:78 were: *Shewanella oneidensis* MR-1 "multi-domain β-ketoacyl synthase" (identity = 67%; positives = 79%); *Shewanella* sp. SCRC-2738 orf5 (identity = 69%; positives = 77%); and *Moritella marina* ORF8 (identity = 56%; positives = 70%). The *Schizochytrium* sp. ATCC20888 orfA was 41% identical and 55% positive to the deduced amino acid sequence encoded by SEQ ID NO:78.

At the amino acid level, the sequences with the greatest degree of homology to the *Shewanella olleyana* ACAM644 dehydrogenase (DH) deduced amino acid sequence encoded by SEQ ID NO:75 were: *Shewanella* sp. SCRC-2738 orf7 (identity = 77%; positives = 86%); *Photobacter profundum* pfaC (identity = 72%; positives = 81%); and *Shewanella oneidensis* MR-1 "multi-domain β-ketoacyl synthase" (identity = 75%; positives = 83%). The *Schizochytrium* sp. ATCC20888 orfC was 26% identical and 42% positive to the deduced amino acid sequence encoded by SEQ ID NO:75.

At the amino acid level, the sequences with the greatest degree of homology to the *Shewanella japonica* ATCC BAA-316 dehydrogenase (DH) deduced amino acid sequence encoded by SEQ ID NO:77 were: *Shewanella* sp. SCRC-2738 orf7 (identity = 77%; positives = 86%); *Photobacter profundum* pfaC (identity = 73%; positives = 83%) and *Shewanella oneidensis* MR-1 "multi-domain β-ketoacyl synthase" (identity = 74%; positives = 81%). The *Schizochytrium* sp. ATCC20888 orfC was 27% identical and 42% positive to the deduced amino acid sequence encoded by SEQ ID NO:77.

It is expected that the PUFA PKS gene sets from these two *Shewanella* strains will provide beneficial sources of whole genes or individual domains for the modification of *Schizochytrium* PUFA production. PUFA PKS genes and the proteins and domains encoded thereby from either of *Shewanella olleyana* or *Shewanella japonica* are explicitly encompassed by the present invention.

Example 8

This example demonstrates how the bacterial PUFA PKS gene fragments described in Example 7 can be used to modify PUFA production in *Schizochytrium*.

All presently-known examples of PUFA PKS genes from bacteria exist as four closely linked genes that contain the same domains as in the three-gene *Schizochytrium* set. It is anticipated that the PUFA PKS genes from *Shewanella olleyana* and *Shewanella japonica* will likewise be found in this tightly clustered arrangement. The homologous regions identified in Example 7 are used to isolate the PUFA PKS gene clusters from clone banks of *Sh. olleyana* and *Sh. japonica* DNAs. Clone banks can be constructed in bacteriophage lambda vectors, cosmid vectors, bacterial artificial chromosome ("BAC") vectors, or by other methods known in the art. Desired clones containing bacterial PUFA PKS genes can be identified by colony or plaque hybridization (as described in Example 1) using probes generated by PCR of the partial gene sequences of Example 7 employing primers designed from these sequences. The complete DNA sequence of the new bacterial PUFA PKS gene sets are then used to design vectors for transformation of *Schizochytrium* strains defective in the endogenous PUFA PKS genes (e.g., see Examples 3, 5, and 6). Whole bacterial genes (coding sequences) may be used to replace whole *Schizochytrium*

genes (coding sequences), thus utilizing the *Schizochytrium* gene expression regions, and the fourth bacterial gene may be targeted to a different location within the genome. Alternatively, individual bacterial PUFA PKS functional domains may be "swapped" or exchanged with the analogous *Schizochytrium* domains by similar techniques of homologous recombination. It is understood that the sequence of the bacterial PUFA PKS genes or domains may have to be modified to accommodate details of *Schizochytrium* codon usage, but this is within the ability of those of skill in the art.

Each publication cited or discussed herein is incorporated herein by reference in its entirety.

While various embodiments of the present invention have been described in detail, it is apparent that modifications and adaptations of those embodiments will occur to those skilled in the art. It is to be expressly understood, however, that such modifications and adaptations are within the scope of the present invention, as set forth in the following claims.